# Enzymatic Properties and Directed Evolution of Transketolase and Chemoenzymatic Synthesis of neo-Sialoconjugates

**by Yi, Dong**

TECHNISCHE
UNIVERSITÄT
DARMSTADT

# Enzymatic Properties and Directed Evolution of Transketolase
# and
# Chemoenzymatic Synthesis of neo-Sialoconjugates

Vom Fachbereich Chemie
der Technischen Universität Darmstadt

zur Erlangung des akademischen Grades eines

Doktor rerum naturalium
(Dr. rer. nat.)

genehmigte
Dissertation

vorgelegt von

**M. Sc. Yi, Dong**
aus Shaoxing, China

| | |
|---|---|
| Referent: | Prof. Dr. Wolf-Dieter Fessner |
| Korreferenten: | Prof. Dr. Katja Schmitz |
| | Prof. Dr. Pere Clapes |
| Tag der Einreichung: | 08.11.2012 |
| Tag der mündlichen Prüfung: | 20.12.2012 |

Darmstadt 2012
**D17**

This work is dedicated to my motherland and my family that have supported me during this period of growth, and to my grandfather who left us too soon.

## Acknowledgements

As a Master's graduate in Shanghai, China, I was very pleased to receive the invitation from Prof. Dr. Wolf-Dieter Fessner to start my Ph.D. in Clements-Schöpf-Institute of Organic Chemistry and Biochemistry of Technische Universität Darmstadt. During my research in Darmstadt, Prof. Fessner has given his great deal of patience for the supervision of my project. No matter how busy he was, he always offered me enough time for our discussion. He has inspired me by his profound knowledge. What I have learnt here is not only the professional knowledge, but also the rigorous attitude and open thinking. I express my heartfelt thanks to him for his genuine instruction.

I also especially thank Prof. Dr. Katja Schmitz for her kind supervision, and Prof. Dr. Harald Kolmar for his full support regarding my DAAD scholarship.

Although Darmstadt is located almost 8,000 kilometers away from my hometown, I could still feel the warmth of home in our research group. I would like to write down everyone's name in our team here to give my gratitude. To start with, I thank Ms. Serbet Pinar-Safi, Mr. Michael Kickstein, Mr. Hans-Michael Orfgen, Ms. Bettina Harnischfeger, Dr. Ning He, Dr. Madhura Rale, Dr. Titu Devamani, Mr. Dirk Heyl, Mr. Thomas Scheidt, Mr. Mark Schnellbächer, Ms. Deniz Güclü, Mr. Sebastian Junker, Ms. Eva Heilig, and Mr. Steffen Kühl.

I thank our cooperative partners, Prof. Dr. Laurence Hecquet's group in Clermont Université, Université Blaise Pascal - France for the transketolase project, Prof. Dr. Alan Berry's team in Leed University - UK for the CSS and sialyltransferase thema, and Prof. Dr. Gerold Barth and Mr. Sebastian Stark in Technische Universität Dresden - Germany for the *Yarrowia* expression system.

I thank German Academic Exchange Service (DAAD) and China Scholarship Council (CSC) for offering me the scholarship to support my research activities and stay in Germany.

Last but not least, I must thank my family for their longtime support and encouragment during this period of growth. They are always my spiritual pillar in my heart.

Yi, Dong
31.10.2012 Darmstadt

## Contents

## Zusammenfassung

### Teil I. Enzymatische Eigenschaften und gelenkte Evolution der Transketolase

- **TK Assay für das Substrat-Fingerprinting**

Transketolase (TK, EC 2.2.1.1) katalysiert den stereospezifischen Transfer einer $C_2$-Ketoleinheit auf das Carbonylende einer Reihe von Aldehyden. Die Verwendung der TK zum Aufbau von Kohlenstoff-Kohlenstoff-Bindungen erhält wegen der hoch stereoselektiven Bildung eines (*S*)-konfigurierten, chiralen Zentrums unter gleichzeitiger Racematspaltung von (2*R*)-Hydroxyaldehyden vermehrte Beachtung. Insbesondere beim Einsatz von Hydroxypyruvat (HPA) als Donor führt die Freisetzung von $CO_2$ zu einer irreversiblen Reaktion wobei das freigesetzte Bicarbonat den pH-Wert des Reaktionssystems erhöht. Aufbauend auf diesem Prinzip wurde, mit HPA als Donor und Phenolrot als empfindlichem pH Indikator, eine neue, kontinuierliche pH-basierte Assaymethode für das Hochdurchsatzscreening der TK entwickelt. Dieser generische Assay kann zuverlässig für die Colorimetrische Bestimmung der spezifischen Enzymaktivität und zur Messung kinetischer Konstanten benutzt werden. Mit dieser Methode wurden sowohl die kinetischen Konstanten und die Akzeptoraffinitäten der TK aus *Escherichia coli* ($TK_{eco}$) und *Saccharomyces cerevisiae* ($TK_{yst}$) als auch zweier $TK_{eco}$-Mutanten, ($TK_{eco}^{D469E}$ und $TK_{eco}^{H26Y}$), für einen detailierten Vergleich untereinander ermittelt. Die quantitativen Daten zeigen, dass das katalytische Potenzial von Wildtyp-$TK_{eco}$ und $TK_{yst}$ sehr ähnlich sind und nur eine geringe Variation aufweisen. Wildtyp-TK-Enzyme haben eine hohe Präferenz für α-hydroxylierte Akzeptoren, zusammen mit einer ausgeprägten Stereoselektivität gegenüber der (2*R*)-Konfiguration. Die Variante $TK_{eco}^{D469E}$ zeigt signifikant höhere Aktivität gegenüber 2-deoxygenierten Aldehyden als das entsprechende Wildtyp Enzym.

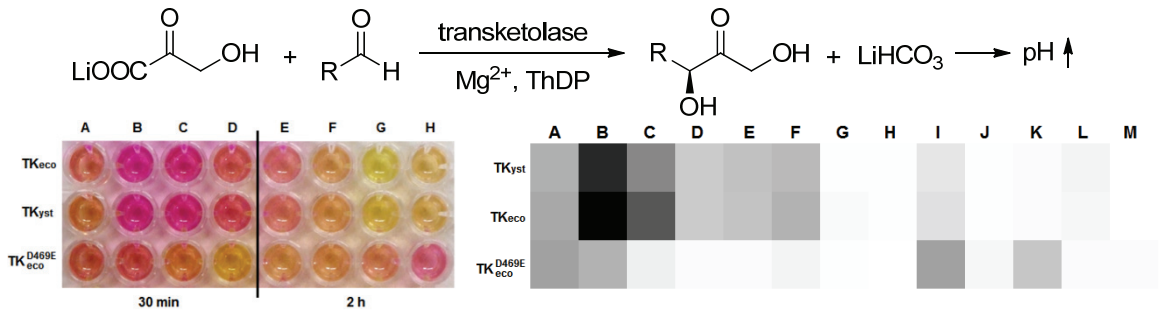



Abbildung 1. Das Prinzip einer neuen, pH-basierten Testmethode für Transketolase und ihre Anwendung zur Bestimmung der spezifischen Enzymaktivität.

- **TK Test auf allgerneine Aldehyd Toleranz**

Die TK aus *Geobacillus stearothermophilus* ($TK_{gst}$) ist ein neues Enzym, das kürzlich kloniert wurde und eine beeindruckende Thermostabilität aufweist. Ähnlich wie andere TKs zeigt es eine geringe Aktivität bei nicht α-hydroxylierten Aldehyden. Um diese Aktivität zu erhöhen wurden in einem ersten Versuch Sättigungsmutagenesebibliotheken der $TK_{gst}$ an den Positionen Leu382 und Asp470 erstellt. Leu382 and Asp470 sind die entsprechenden Aminosäurereste zur Bindung der Hydroxymethylen-Gruppe von α-hydroxylierten Aldehyden. Das Screening der Bibliothek lieferte mehrere positive Varianten, sowohl aus der einfach, als auch aus der doppelt mutierten Bibliothek. Als herausragende Mutante konnte die Variante Asp469Ile mit einer 17-, 13- und 8.6-fachen Aktivitätsverbesserung im Vergleich zum Wildtyp-Enzym gegenüber Propanal, Butanal and Methoxyethanal identifiziert werden. Ihre hohe Thermostabilität mit einer $T_m$ von 74 °C bleibt erhalten.

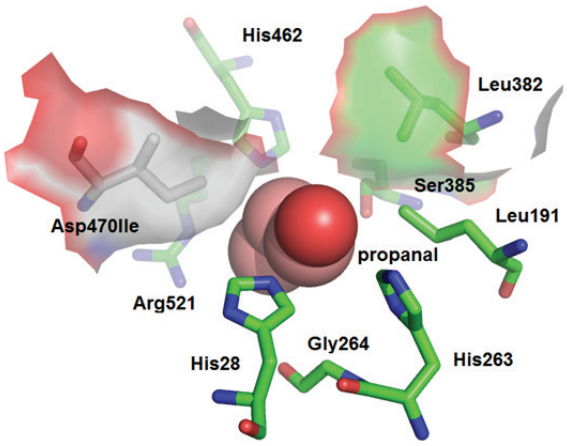



Abbildung 2. Kupplung von Propanal als Substratanalog für $TK_{gst}^{D469I}$. Asp470Ile und Leu382 bilden eine unpolare Oberfläche um die aliphatische Kette von Propanal mit höherer Affinität zu binden.

## Teil II. Chemoenzymatische Synthese von Neosialokonjugaten

- **SiaT Assay für die Enzymkinetik**

Sialinsäuren sind Kohlenhydrat-Derivate bestehend aus neun Kohlenstoffatomen die eine Carboxylaruppe besitzen. Sie kommen weit verbreitet frei oder konjungiert an Glycanen der Zelloberfläche von und erfüllen dort wichitge Funktionen in der Zellphysiologie. Deshalb ist es wichtig, Methoden für die Synthese spezifisch sialokonjugierter Epitope *in vitro* zu entwickeln, um die Forschung von Kohlenydrat-Protein-Interaktionen in physiologischen Prozessen zu ermöglichen. Sialinsäurereste werden von α/β-Galactosyl-α-2,3/2,6-

Sialyltransferasen auf die terminale Galactoseeinheit von Oligosacchariden übertragen. Deshalb stellen diese nützliche Werkzeuge für die Synthese von Sialokonjugaten *in vitro* dar. Zu diesem Zweck wurden α-2,3- und 2,6-Sialyltransferasen aus *Photobacterium phosphoreum* (2,3SiaT$_{pph}$) und entsprechend aus *Photobacterium leiognathi* JT-SHIZ-145 (2,6SiaT$_{ple}$) kloniert. Um die Substratspezifität dieser beiden SiaTs zu untersuchen, wurde zur Bestimmung der kinetischen Konstanten und der Substrattoleranz ein pH-basierter Assay etabliert, in dem Phenolrot als pH-Indikator in schwach gepuffertem Medium dient. Diese pH-Assay-Methode ist schnell, empfindlich und kontinuierlich und konnte zur Bestimmung der kinetischen Konstanten und der Substrattoleranz von 2,3SiaT$_{pph}$ und 2,6SiaT$_{ple}$ herangezogen werden. Diese beiden SiaTs weisen eine deutlich unterschiedliche Akzeptorspezifität auf. Während die 2,3SiaT$_{pph}$ α- und β-Galactoside, Lactoside, Glc, Man und GalNAc gut akzeptieren kann, zeigt die 2,6SiaT$_{ple}$ nur gegenüber β-Galactosiden und Lactosiden eine hohe Aktivität.

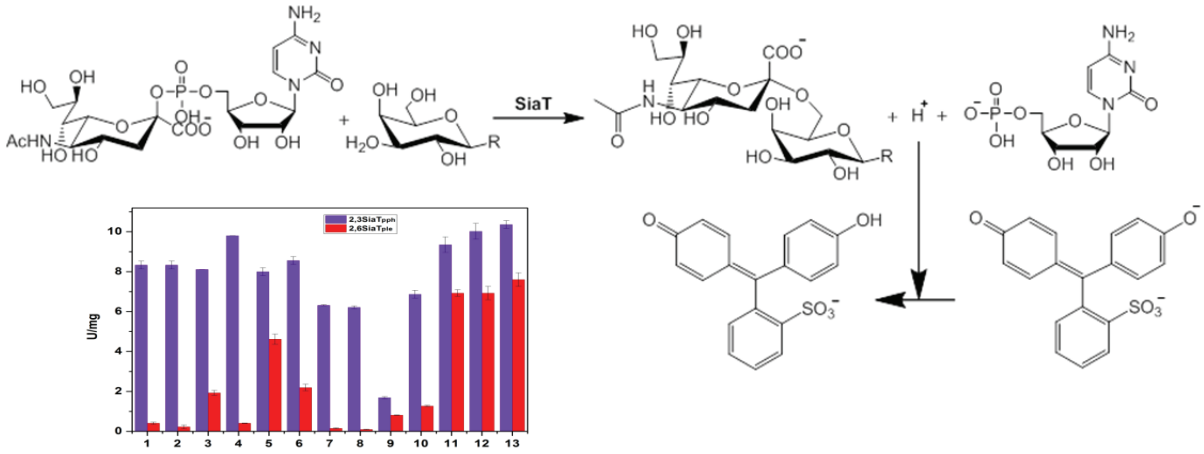



Abbildung 3. Das Prinzip einer neuen, pH -basierten Testmethode für Sialyltransferase und ihre Anwendung zur Bestimmung der spezifischen Enzymaktivität.

- **β-SiaT Engineering für α-Substrattoeranz**

Durch den Strukturabgleich zwischen 2,3SiaT$_{pph}$ and 2,6SiaT$_{ple}$ wurde Trp347 als Ursache für die Blockade des aktiven Zentrums für α-Galactoside identifiziert. Um die Aktivität der 2,6SiaT$_{ple}$ bezüglich α-Galactosiden zu verbessern wurden drei Mutagenesebibliotheken erstellt. Hierfür wurde die Hochdurchsatz-Screeningmethode mit Methyl-α-D-galactopyranosid als Akzeptor verwendet. Es konnte jedoch keine positive Variante identifiziert werden, was zeigt, dass Trp347 in der 2,6SiaT$_{ple}$ offenbar nicht problemlos ersetzbar ist. Es ist zu vermuten, dass Trp347 zur Bindung zwischen Donor und Akzeptor beiträgt und die korrekte Akzeptor-Konformation für die Bildung von 2,6-sialisierten Produkten stabilisiert.

- **CSS Engineering für sterisch auspruchsvolle Substrate**

Die CMP-Sialinsäure-Synthetase (CSS, E.C. 2.7.7.43) katalysiert die CMP-Aktivierung von Sialinsäuren unter Verwendung von CTP als Donor und $Mg^{2+}$ als Cofaktor. Es konnte gezeigt werden, dass eine Reihe von Neu5Ac-Analogen mit einer sterisch auspruchsvollen Modifikation an der Acylamino-Gruppe von CSS aus *Neisseria meningitidis* ($CSS_{nme}$) akzeptiert werden können, obwohl ihre korrespondierenden $k_{cat}/K_M$ Werte viel niedriger sind als diejenigen des natürlichen Substrates Neu5Ac. Durch ein Alignment des aktiven Zentrums für die Akzeptorbindung der $CSS_{nme}$ wurden Phe192 und Phe193 als die entsprecheuden Reste für die Bindung der Acylaminogruppe in Neu5Ac identifiziert. Darauf aufbauend wurden vier Sättigungsmutagenesebibliotheken an den Positionen Phe192 und Phe193 konstruiert, um die Aktivität gegen Neu5Ac-OBz, Neu5Hex und $Neu5PenN_3$ zu verbessern. Aus dem Screening mit der pH-basierten Hochdurchsatzmethode ergaben sich $CSS_{nme}$-Phe192Cys/Phe193Tyr, $CSS_{nme}$-Phe192Ser, und $CSS_{nme}$-Phe192Ser/Phe193Tyr als die besten Kandidaten mit jeweils 14.1-, 4.6- und 4.0-facher Aktivitätssteigerung gegen Neu5Ac-OBz, Neu5Hex und $Neu5PenN_3$. Um die verbesserte Aktivität zu validieren, wurden diese drei CSS-Varianten zusammen mit der $2,6SiaT_{ple}$ für eine präparative Synthese von neuen Neo-Sialokonjugaten verwendet. Die Ausbeuten der Endprodukte betrugen 50%, 64% bzw. 70%.

- **Enzymklonierung für die Synthese von LacNAc**

Oligosacharide mit einer endständigen *N*-Acetyl-D-lactosamin Einheite (LacNAc) sind ein wichtiges natürliches Substrat der SiaTs. Die Synthese von LacNAc und seiner Derivate *in vitro* erweitert daher die Akzeptor Vielfalt für die Herstellung von Neo-Sialokonjugaten. Für diesen Zureck wurden die β-1,4-Galactosyltransferase aus *Helicobacter pylori* ($GalT_{hpy}$) und die UDP-Galactose-4-Epimerase aus *E. Coli* ($GalE_{eco}$) kloniert und in einem *E. coli*-Expressionssystem exprimiert. Mit Hilfe dieser beiden rekombinanten Enzyme wurden freies LacNAc und LacNAc-D-T-P-Acr aus UDP-Glc als Donor sowie den entsprechenden Akzeptorsubstraten in isolierten Ausbeuten von 88%, 92% und 85% jeweils im 100 mg-Maßstab für weitere präparative Zwecke synthetisiert.

- **Präparative Synthese von Neo-Sialokonjugaten**

Unter Verwendung der $2,3SiaT_{pph}$ bzw. $2,6SiaT_{ple}$ sowie der neuen CSS-Varianten wurden die Neo-2,3- und 2,6-Sialokonjugate in einer Eintopf-Reaktion mit zwei Enzymen synthetisiert.

Sechs typische Sialinsäuren (Neu5Ac, Neu5Gc, Kdn, *epi*-Kdn, Neu5Ac-9-$N_3$ und Neu5NPhAc) mit Modifikationen an C5 und C9 wurden als Donoren an fluoreszente Lactose (Lac-D-T-P-Acr) und *N*-Acetyl-D-lactosamin (LacNAc-D-T-P-Acr) mit α-2,3- and α-2,6-Konfiguration gekuppelt. Die 2,3SiaT$_{pph}$ und 2,6SiaT$_{ple}$ zeigen sehr ähnliche Donorspezifitäten gegen über diesen sechs Sialinsäuren. Beide akzeptieren Neu5Ac, Neu5Gc, Kdn und Neu5Ac-9-$N_3$ gut, reagieren jedoch mit dem sterisch auspruchsvollen Neu5NPhAc schlechter. Darüber hinaus kann die 2,6SiaT$_{ple}$ das *epi*-Kdn besser als die 2,3SiaT$_{pph}$ akzeptieren.

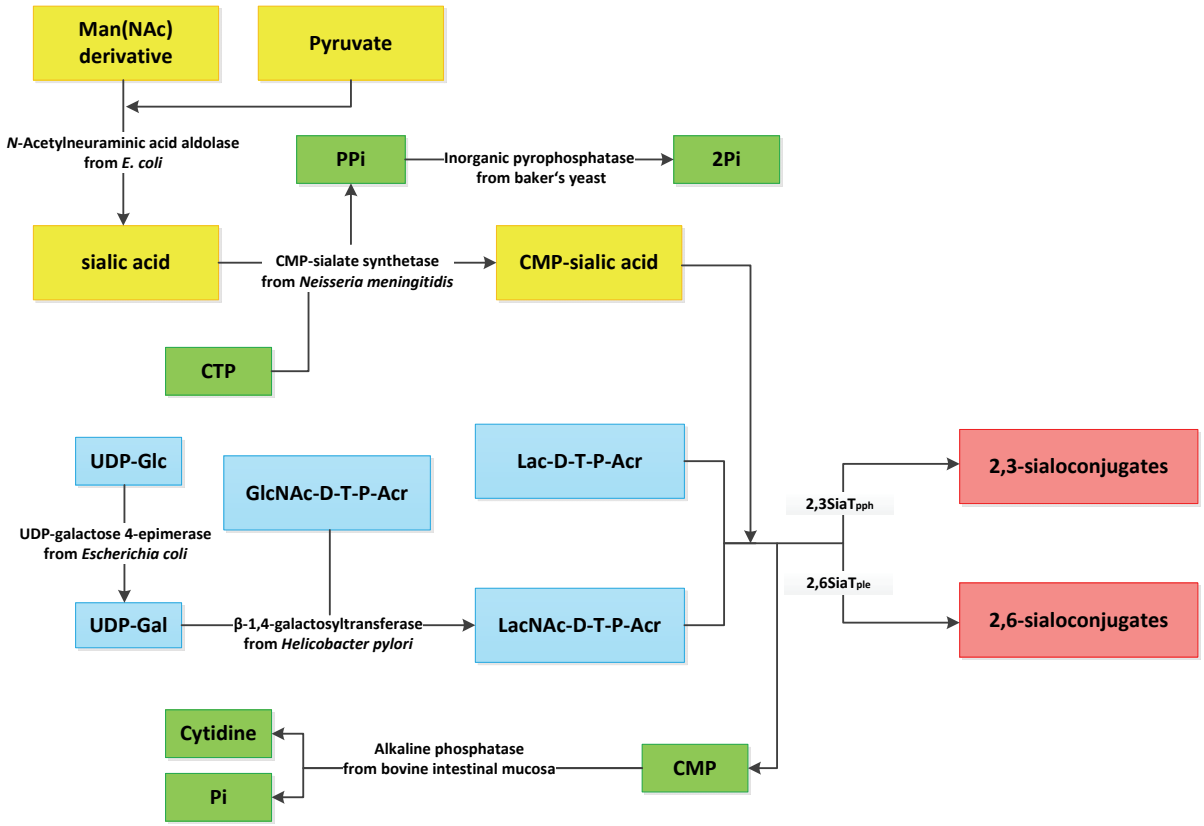



Abbildung 4. Chemoenzymatische Syntheseroute für die fluorogenen Sialokonjugate.

- **Charakterisierung einerpotentiellen neuen SiaT**

Kürzlich wurde eine neue potentielle Sialyltransferase aus *Shewanella piezotolerans* (SiaT$_{spi}$) beruhend auf der genomischen DNA-Sequenz von *S. piezotolerans* beschrieben. Über eine Klonierung, Expression und einen Aktivitätstest der SiaT$_{spi}$ liegen jedoch noch keine Berichte vor. Deshalb wurde das künstliche Gen im Rahmen dieser Arbeit kloniert, modifiziert und optimal im *E. coli* und *Yarrowia*-Expressionssystem exprimiert. Der Sequenz-BLAST und die strukturelle Übereinstimmung mit anderen 2,3SiaT-Sequenzen innerhalb der CAZy-Familie GT80 ließen eine funktionelle upstream-Sequenz erkennen. Die Deletion eines Loops mit 12

Aminosäurenresten verbesserte die Identifizierung der 2,3-SiaT-Aktivität durch eine Reaktion mit fluoreszenzmarkiertem Akzeptor, insbesondere bei Koexpression mit Chaperonen. Die geringe Expression an löslichem Protein sowohl im *E. coli*-, als auch im *Y. lipolytica*-Expressionssystem vereitelte jedoch eine weitergehende Charakterisieng des Enzymes.

## Abbreviation

| | |
|---|---|
| 2,3SiaT | α/β-galactoside α-2,3-sialyltransferase |
| $2,3SiaT_{pph}$ | α/β-galactoside α-2,3-sialyltransferase from *Photobacterium phosphoreum* |
| 2,6SiaT | β-galactoside α-2,6-sialyltransferase |
| $2,6SiaT_{pda}$ | 2,6SiaT from *Photobacterium damselae* JT0160 |
| $2,6SiaT_{ple}$ | 2,6SiaT from *Photobacterium leiognathi* JT-SHIZ-145 |
| $2,6SiaT_{psp}$ | 2,6SiaT from *Photobacterium* sp. JT-ISH-224 |
| AA | amino acid |
| Ac | acetyl |
| AcOH | acetic acid |
| Acridon/Acr | 10*H*-acridin-9-OH |
| Amp | ampicillin |
| Bn | benzyl |
| BuOH | butanol |
| CAZy | The Carbohydrate-Active Enzymes Database |
| CDP | cytidine diphosphate |
| CMP | cytidine monophosphate |
| CSS | CMP-Neu5Ac synthetase |
| $CSS_{mmu}$ | CSS from *Mus musculus* |
| $CSS_{nme}$ | CSS from *Neisseria meningitidis* |
| CTP | cytidine triphosphate |
| $ddH_2O$ | double-distilled water |
| DHAP | triose phosphates dihydroxyacetone phosphate |
| DHEThDP | α,β-dihydroxyethyl ThDP |
| E4P | D-erythrose 4-phosphate |
| EC | Enzyme Commission |
| EtOH | ethanol |
| F6P | D-fructose-6-phosphate |
| FruA | fructose-1,6-bisphosphate aldolase |
| $FucT_{hpy}$ | fucosyltransferase from *Helicobacter pylori* |
| G3P/GAP | D-glyceraldehyde-3-phosphate |

| | |
|---|---|
| Gal | galactose |
| GalE | UDP-galactose 4-epimerase |
| $GalE_{eco}$ | UDP-galactose 4-epimerase from *E. coli* |
| GalNAc | *N*-acetylgalactosamine |
| GalT | β-1,4-galactosyltransferase |
| $GalT_{hpy}$ | β-1,4-galactosyltransferase from *Helicobacter pylori* |
| $GalT_{ngo}$ | β-1,4-galactosyltransferase from *Neisseria gonorrhoeae* |
| $GalT_{nme}$ | β-1,4-galactosyltransferase from *Neisseria meningitides* |
| GalU | UTP-glucose-1-phosphate uridylyltransferase |
| Glc | glucose |
| GlcNAc | *N*-acetylglucosamine |
| GNE | UDP-GlcNAc 2-epimerase/ManNAc-6-kinase |
| GT-B | glycosyltransferase-B |
| HPA | hydroxypyruvate |
| HPLC | high-performance liquid chromatography |
| IPTG | isopropyl β-D-1-thiogalactopyranoside |
| ISM | Iterative Saturation Mutagenesis |
| Kan | kanamycin |
| $k_{cat}$ | Michaelis-Menten equation |
| Kdn | deaminoneuraminic acid |
| $K_M$ | Michaelis constant |
| Lac | lactose |
| LacNAc | *N*-acetyl-D-lactosamine |
| LB | Lysogeny Broth |
| LOD | limit of detection |
| LOQ | limit of quantification |
| Man | mannose |
| ManNAc | *N*-acetyl-D-mannosamine |
| ManNAc-6-P | *N*-acetyl-D-mannosamine-6-phosphate |
| MBP | maltose binding protein |
| MeOH | methanol |
| $NAD^+$ | nicotinamide adenine dinucleotide |
| NADH | reduced nicotinamide adenine dinucleotide |
| NANP | Neu5Ac-9-phosphate phosphatase |
| NANS | Neu5Ac-9-phosphate synthase |

| | |
|---|---|
| Neu | neuraminic acid |
| Neu5Ac | *N*-acetylneuraminic acid |
| Neu5Gc | *N*-glycolylneuraminic acid |
| NeuC | UDP-GlcNAc 2-epimerase |
| NeuS | Neu5Ac synthase |
| NMR | nuclear magnetic resonance |
| PCR | polymerase chain reaction |
| PDB | The Protein Data Bank |
| PEP | phosphoenolpyruvate |
| PGA | penicillin G acylase |
| PGMs | post-glycosylation modifications |
| PhAc | phenylacetyl |
| $P_i$/P | phosphate |
| p$K_a$ | acid dissociation constant |
| *p*NP-Man | 4-nitrophenyl labeled mannose |
| PPase | inorganic pyrophosphatase |
| $PP_i$ | pyrophosphoric acid |
| R5P | D-ribose 5-phostphate |
| R5P | D-ribose-5-phostphate |
| $R_f$ | retardation factor |
| S7P | D-sedoheptulose-7-phosphate |
| SDS-PAGE | sodium dodecyl sulfate polyacrylamide gel electrophoresis |
| SiaT | sialyltransferase |
| $SiaT_{spi}$ | SiaT from *Shewanella piezotolerans* |
| $SL^X$ | Sialyl Lewis$^X$ |
| TB | Terrific Broth |
| TEA | triethanolamine |
| TEED | The Thiamine Diphosphate Dependent Enzyme Engineering Database |
| ThDP | thiamine diphosphate |
| THF | tetrahydrofuran |
| TK | transketolase |
| $TK_{ban}$ | transketolase from *Bacillus anthracis* |
| $TK_{eco}$ | transketolase from *E. coli* |
| $TK_{gst}$ | transketolase from *Geobacillus stearothermophilus* |
| $TK_{yst}$ | transketolase from baker's yeast |
| TLC | thin layer chromatography |

| | |
|---|---|
| Tris | 2-amino-2-hydroxymethyl-propane-1,3-diol |
| UDP | uridine triphosphate |
| UV | ultraviolet |
| V | volumn |
| $V_{max}$ | the maximum reaction rate |
| X5P | D-xylulose 5-phosphate |
| X-gal | 5-bromo-4-chloro-indolyl-β-D-galactopyranoside |

## 1. General Introduction——White Biotechnology and Its Research Progress

White biotechnology, also known as industrial biotechnology, is a subject which applies the tool of nature for the manufacture of chemical products.[1, 2] In the vast majority of cases, bioactive molecules (mainly isolated enzymes) and/or whole cells are utilized as biocatalysts in the production of valuable intermediates or secondary metabolites as final products. The application of white biotechnology to the synthesis of organic fine chemicals and chiral intermediates also led to a “green chemistry” revolution in the chemical industry.[3-5] Because of the biocatalysts’ high catalytic efficiency, substrate specificity and product stereoselectivity, as well as its environment friendliness, the application of biocatalytic process is creating a trend to replace conventional chemical synthesis in pharmaceutical,[6-8] fine chemical,[8, 9] material,[10-12] energy,[12, 13] and environment protection[11, 12] industries, in order to increase the quality and consistency of products, save energy consumption and reduce environmental pollution.[14]

Biocatalysis has experienced a rapid development in the past century, during which many new technologies have been applied for broadening the use of biocatalysts. The application of thermophilic enzymes and immobilization greatly improved the stability of biocatalysts for synthesis.[15, 16] Coenzyme regeneration technology and whole cell catalysis ensure the sustainability and decrease the cost of biocatalytic reactions.[17] Biocatalysis in organic solvents changes the behavior of the enzymes and extend the scope for their application.[17-20] Multiphase reaction systems are advantageous to improve the catalytic efficiency and facilitate product separation.[21] The improvement of enzyme reactors, especially the construction of microreactors, helps to optimize the enzymatic reaction conditions.[22, 23] Owing to the technology above, the substrates and products of biocatalysis are no longer confined to the natural molecule diversity, but also applicable to many non-natural target compounds which are of interest as pharmaceutical intermediates and fine chemicals of interest.[24]

The discovery of suitable or improved biocatalysts is the main theme in biocatalysis. Screening tens of thousands of microorganisms in the environment is a traditional but currently still important and indispensable method. However, with a view to numerous nonculturable microorganisms, the traditional cultivation and screening method cannot meet the

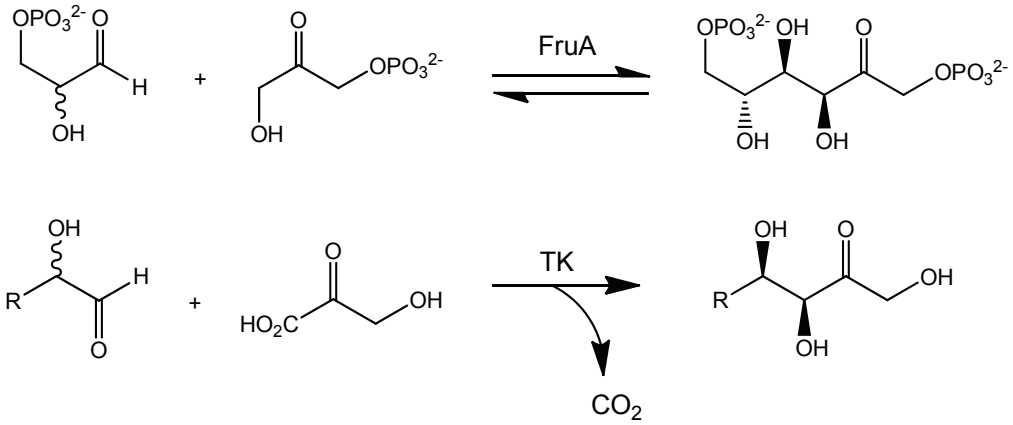
requirements from an increasing demand for new and better catalysts. Fortunately, metagenomic technology enables us to exploit the genetic database of microbial communities without any cultivation step.[25] Based on the bioinformatic analysis, potential isozymes can be identified relatively easy by sequence alignment. In the past decades, new DNA technologies, especially low cost but high-throughput DNA sequencing and artificial synthesis,[24] and new bioinformatic arithmetics have been well developed and paved the way for metagenomics, which develops to become a hot topic in the field of biocatalysis.[25]

On the other hand, protein engineering has created new ways for the modification of the physical and chemical properties of existing biocatalysts.[24] Based on the structural analysis of known enzymes, and through the rational or non-rational design of variants together with high throughput screening methods for the desired function, directed evolution becomes an efficient approach to obtain novel enzyme variants with improved specific activity, substrate specificity, stereoselectivity, thermostability and organic solvent tolerance, or even new catalytic capability. Protein crystal structure analysis with substrate docking simulation supports the design of modified structure and function and prediction of variants of interest, which ultimate provides the theoretical basis for the rational protein design, making *de novo* design of novel enzymes possible.[24] Owing to the combination of rational design and directed evolution, focused cassette and saturation mutagenesis can immensely reduce the size of a mutagenesis library and improve the efficiency of forward mutation in comparison with the irrational approaches.[26] All of these technological developments in biotechnology and bioinformatics have accelerated to a third wave in biocatalysis.[24]

After many years of technological development, a vast diversity of all kinds of enzyme classes have been successfully used as biocatalysts in organic synthesis, including oxidoreductases, transferases, hydrolases, lyases, isomerases and ligases.[5, 14, 27-29] Among them, enzymatic carboligation and glycosylation have become very important reactions for the synthesis of monosaccharides, oligosaccharides and their derivatives *in vitro*.

Several enzymes are involved in the enzymatic carboligation, which can rapidly build a new C-C bond up under mild conditions with high chemical efficiency and uncompromised stereochemical fidelity, without a need for protection of sensitive or reactive functional groups.[30] Aldolase is such kind of biocatalyst containing various members. Fructose-1,6-bisphosphate aldolase (FruA, EC 4.1.2.13) is a well-known aldolase, which transfers the aldol and fructose 1,6-bisphosphate into the triose phosphates dihydroxyacetone phosphate (DHAP) and glyceraldehyde 3-phosphate (GAP).[29] However, the reversibility of the reaction limits the maximum yield of the final product. Transketolase (TK, EC 2.2.1.1) is another essential

enzyme in carbohydrate metabolism. It transfers a terminal two-carbon hydroxyacetyl unit from a donor ketose phosphate to an acceptor aldose phosphate.[31] Although in principle this is also a reversible enzymatic reaction, the utilization of hydoxypruvate as an alternative donor analog can release carbon dioxide during the reaction, which recedes the transfer irreversible.[32] Therefore, TK is an emerging catalyst for asymmetric carbon-carbon bond formation and sugar-like compound synthesis.[30] The first part of this thesis focuses on the development of TK for preparatic asymmetric synthesis, including the study of its substrate specificity, kinetics, and directed evolution of substrate tolerance towards non-α-hydroxylated substrates.




Scheme 1.1. Carboligation catalyzed by fructose-1,6-bisphosphate aldolase and transketolase

Glycosyltransferases are catalyzing in the glycosylation reactions in the assembly of oligosaccharides. The monosaccharide donor sugars are mostly required in the activated form of nucleoside diphosphate or monophosphate sugars, lipid phosphates, and free anomeric phosphates.[33] Sialyltransferases are one kind of nucleoside monophosphate sugar dependent glycosyltransferases that transfer CMP-activated sialic acids onto galactosides, or to other sialic acids, forming an α-2,3, α-2,6 or α-2,8 glycosidic bond, respectively.[34] The sialic acid terminus in oligosaccharides plays a very important role in cell information display and exchange in the physical activities, such as cellular adhesiveness and recognition, blood components and clotting, immunology and infection, transmission of nerve impulses, hormonal regulation, and so on.[34, 35] Therefore, the chemoenzymatic synthesis of sialoconjugates *in vitro* has significance for the interpretation of physiological functions of specific oligosaccharide structures and for potential drug design. In part II of this thesis, bacterial sialyltransferases are studied for the synthesis of neo-sialoconjugates. Galactosyltransferase, UDP-glucose 4-epimerase and cytosine-5'-monophosphate-*N*-

acetylneuraminate synthetase are also included owing to their involvement in the acceptor synthesis and donor activation.

# Part I. Enzymatic Properties and Directed Evolution of Transketolases

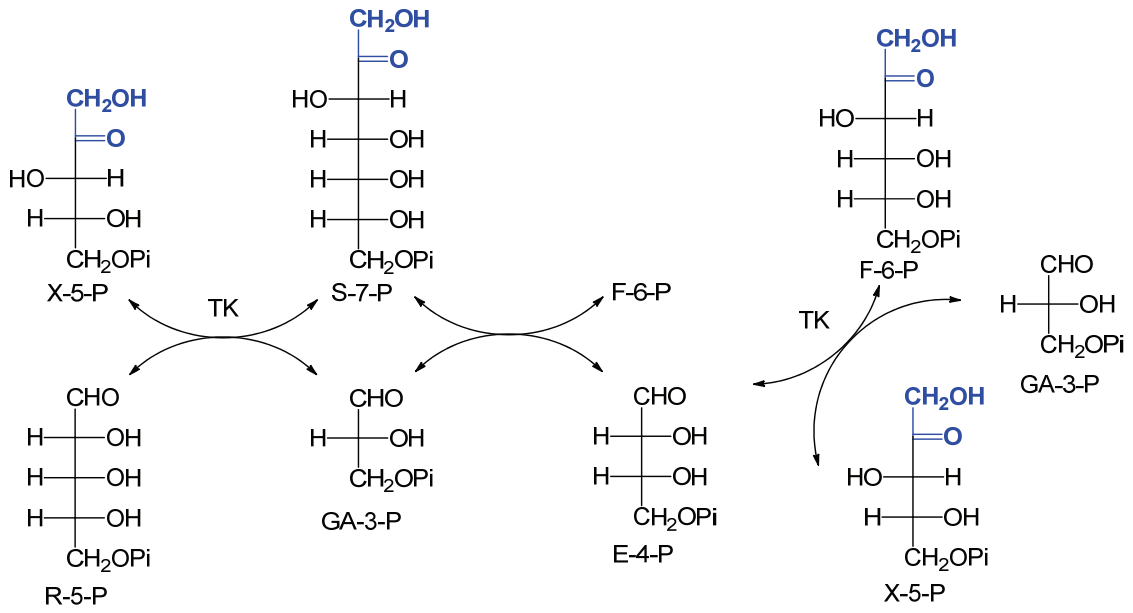
# 2. Introduction of Part I

## 2.1. ThDP-dependent enzymes

Thiamine diphosphate (ThDP) -dependent enzymes are a large family of natural biocatalysts with ThDP as an essential cofactor that catalyze the clearage and formation of carbon-sulfur, carbon-oxygen, carbon-nitrogen and carbon-carbon bonds.[36, 37] A data base named “The Thiamine Diphosphate Dependent Enzyme Engineering Database” (TEED) has been established including the identified ThDP-dependent enzymes as well as potential sequences.[38] Up to the present, more than 9,000 ThDP-dependant enzymes have been identified and additional 3,000 analogous sequences are listed in GenBank.[38] According to their structural arrangement of domains, ThDP-dependent enzymes are classified into six superfamilies, including DC (decarboxylase), TK (transketolase), OR (oxidoreductase), KDH (α-ketoglutarate dehydrogenase), and two subfamilies K1 and K2 of the KD (2-ketoacid dehydrogenase). In each category, the amino acid residue sequences of the enzymes have highly conserved domains which are in charge of the ThDP and functional group binding.[38] Due to their intrinsic chiral function, ThDP-dependent enzymes have been broadly applied in enzymatic synthesis, especially for carbon-carbon bond formation.[39-43]

## 2.2. Transketolase

Transketolase (EC 2.2.1.1, TK) is a ThDP-dependent enzyme performing C-C bond formation and cleavage.[42] It plays an important role in both the pentose phosphate pathway and the Calvin cycle of photosynthesis *in vivo*, catalyzing a reversible 2-carbon fragment transfer reaction between an aldose and a ketose. In the pentose phosphate pathway, it transfers the 2-carbon fragment from D-xylulose 5-phosphate (X5P) to D-ribose 5-phostphate (R5P), and also from X5P to D-erythrose 4-phosphate (E4P). In the Calvin cycle of photosynthesis, it operates in an opposite direction (Scheme 2.1).



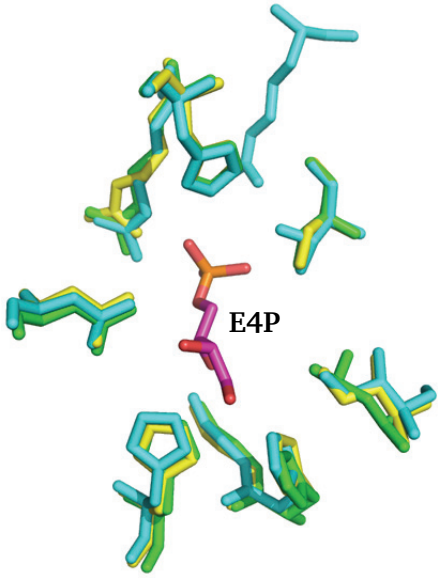

Scheme 2.1. Transketolase function in the pentose phosphate pathway and the Calvin cycle of photosynthesis

As a key enzyme in metabolism, TK has been found widely in both eukaryotes and prokaryotes, including animal, plant, yeast, bacteria, and so on.[42, 44] Among them, bacterial TK have particularly high advantages for chemoenzymatic synthesis, such as high soluble expression in prokaryotic expression systems, relatively high stability and broad substrate range.[42] The most extensively studied bacterial TKs are those from *E. coli* ($TK_{eco}$)[45] and yeast ($TK_{yst}$)[46]. The substrate specificity determination of $TK_{eco}$ and $TK_{yst}$ shows that both of them also have a wide acceptor tolerance for non-phosphorylated aldose sugers.[46-49] When the carbon chain of aldoses is longer than four, the activity of TK decreases very fast, mainly because of the intramolecular formation of cyclic hemiacetals. On the other hand, TK has very strict enantioselectivity towards the α-position of the aldehyde acceptors, for which a hydroxyl group in D-configuration is required for activity. Deoxy-aldedydes or L-α-hydroxylated aldehydes, such as propanal and L-glyceraldehyde, respectively, cannot be well accepted by TK.[46-49]

The TKs share common domains with each other, which contain a pyrophosphate and a pyrimidine domain, as well as the TK C-terminal domain.[38] From the determination of the crystal structures of $TK_{eco}$ (PDB code 2R8O and 2R8P),[50, 51] $TK_{yst}$ (PDB code 1GPU)[52] and TK from *Bacillus anthracis* ($TK_{ban}$, PDB code 3M49), it became apparent that these three bacterial TKs have almost the same active site arrangement (Figure 2.1), which indicates that

$TK_{eco}$, $TK_{yst}$ and $TK_{ban}$ have the same catalytic mechanism for their ketol transfer reactions, because of the highly conserved protein residues that form their active sites. The key residues were identified as Glu418 and His481 (numbers refer to the $TK_{yst}$ sequence), in which the former forms a hydrogen bond to N1' of the aminopyrimidine ring of ThDP, while the latter is positioned to abstract the proton from the active imino group.[42]

a)




b)

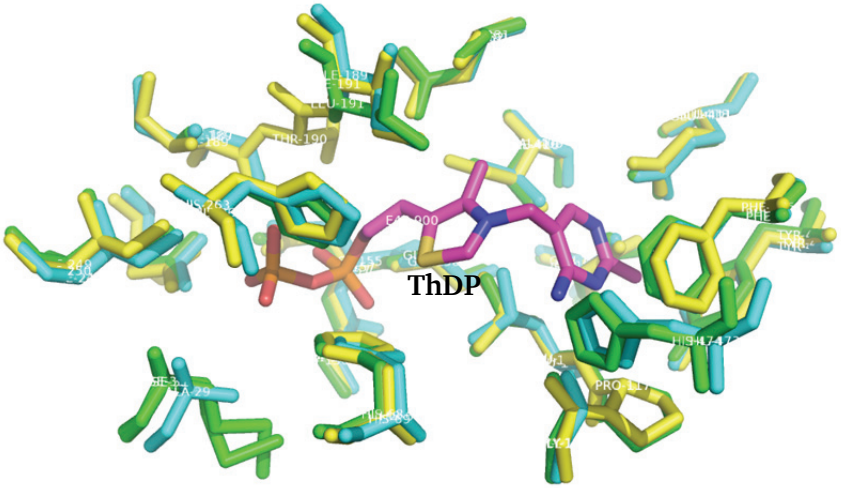
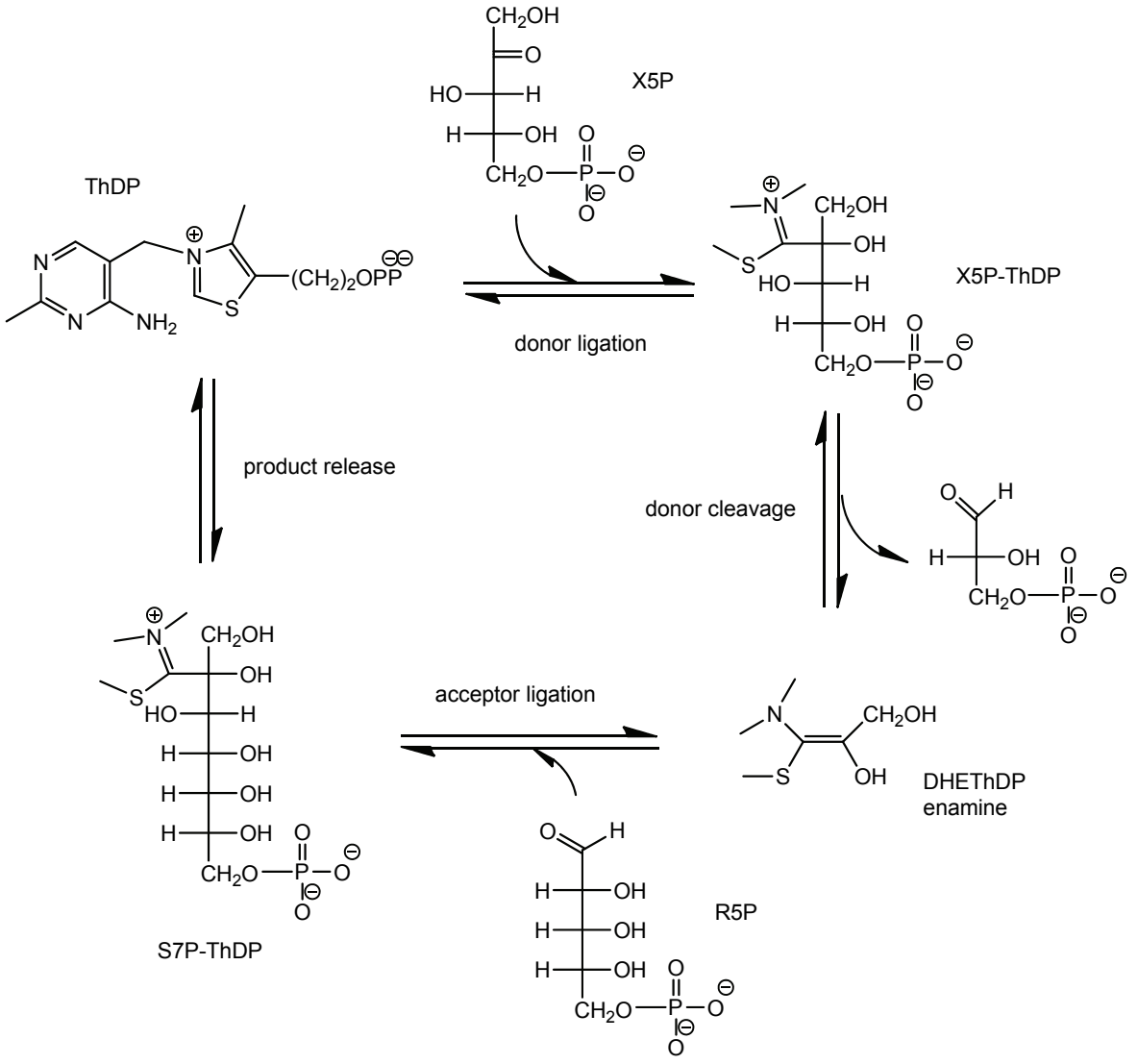



Figure 2.1. Active site alignment of $TK_{eco}$ (blue, PDB code 2R8O and 2R8P), $TK_{yst}$ (yellow, PDB code 1GPU) and $TK_{ban}$ (green, PDB code 3M49). a) Active psite alignment for substrate binding D-erythrose 4-phosphate (E4P). b) Active site alignment for ThDP cofactor binding.

The kinetic analysis of $TK_{eco}$ indicated that the catalysis of TK is consistent with a Ping Pong Bi Bi mechanism, where the ketol donor is bound to the enzyme first, followed by the binding of the aldehyde acceptor and subsequent release of the product.[53] The catalytic mechanism of

this ketol transfer reaction has been elucidated in detail in the resent literature.[51, 54-57] The catalytic reaction starts by the deprotonation of ThDP at the thiazolium ring to form a carbanion. Then, this carbanion binds to the carbonyl of the ketose donor substrate (e.g. X5P) followed by the cleavage of the C2-C3 bond. After the release of the truncated donor fragment (e.g. G3P), ThDP carries the covalent bonding keto fragment affording an α,β-dihydroxyethyl ThDP (DHEThDP) enamine. When an acceptor substrate (e.g. R5P) enters the active site and binds to the DHEThDP enamine, the ketol fragment is transferred to the acceptor forming the product (e.g. S7P), which is subsequently released. Because the product also contains a ketol group, it can also serve as a ketol donor, which makes this reaction fully reversible (Scheme 2.2).




Scheme 2.2. Mechanism of ketol transfer reaction catalyzed by TK.[51]

## 2.3. Directed evolution of transketolase

Although TK can accept non-phosphorylated suger substrates, the activity of TK is much lower than towards natural phosphorylated acceptors. In order to improve its activity, saturation mutagenesis libraries have been constructed on the active sites of $TK_{eco}$ and $TK_{yst}$.[58-60] Twenty amino acid residues surrounding ThDP and acceptor have been chosen as the candidates for mutation.[58] The top variant $TK_{eco}^{H461S}$ shows a 4.8 fold activity improvement towards glycoaldehyde when using hydroxypyruvate (HPA) as donor.[58] On the other hand, owing to the strict enantioselectivity of TK towards the α-hydroxyl group of its acceptors, the saturation mutagenesis libraries have also been screened for extending the substrate range of TK towards non-α-hydroxylated aldehydes which led to a variant $TK_{eco}^{D469E}$ with eight fold activity increase and improved stereoselectivity (doubling the *ee*) for propanal as compared to the wild-type.[59, 61] Moreover, one single-point mutation ($TK_{eco}^{H26Y}$) was found to reverse the stereoselectivity of $TK_{eco}$ towards propanal, thereby forming a product with 3*R* configuration preferentially.[61]

## 2.4. Synthetic application of transketolase

TK performs the transfer of a two-carbon hydroxyacetyl unit to an aldehyde with carbon-carbon bond formation and induction of a chiral hydroxyl methylene group at the C3 position of the product. This makes TK a potent biocatalyst for chiral synthesis, especially in the synthesis of ketose-type intermediates, which is difficult to achieve by chemical methods.[39] However, a major challenge for applications in practical synthesis is that the TK reaction is reversible, which leads to an equilibrium between substrates and products, requires tedious separation, and limits the product yields. To deal with this problem, hydroxypyruvate (HPA) was introduced as a very suitable donor that forms $CO_2$ after the two-carbon unit transformation.[62] The release of $CO_2$ makes the reaction essentially non-reversible, and the reaction can be completed totally. With HPA as donor, a series of chiral polyhydroxyketones have been synthesized including, but not limited to, monosaccharide derivatives and deoxyketose-type products.[46-48, 63, 64]

## 2.5. Specific aim

The substrate specificity and kinetic constants of TK enzymes from diverse resources have been previously determined independently by a variety of assay methods, which makes a direct comparison of their substrate preferences highly inaccurate. Therefore, a rapid, sensitive, continuous and universal assay method for TK needs to be established first for the activity comparison between diverse TKs, as well as for a facile screening of mutation libraries. Based on such a "universal" method, the kinetic constants and acceptor specificity of $TK_{eco}$ and $TK_{yst}$ could be determined for a detailed comparison among each other.

Recently, a thermostable TK from *Geobacillus stearothermophilus* ($TK_{gst}$) has been cloned and expressed in an *E. coli* expression system.[65] Similar to other TKs, it also has low activity towards non-α-hydroxylated aldehydes. In order to improve its activity, *in vitro* mutagenesis of $TK_{gst}$ promises a rapid route to novel variants with advanced properties.

# 3. Study of the Enzymatic Properties of Transketolase Using a pH-Based Assay Method

## 3.1. Introduction

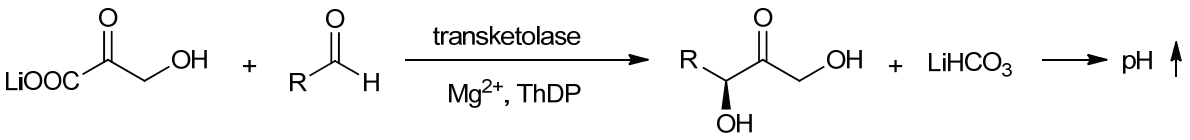
The enzymatic properties, especially the substrate specificity of bacterial TKs including towards phosphorylated and non-phosphorylated acceptor aldehydes, have been well determined in several studies.[39, 42] Unfortunately, most of the specificity data of TKs from various resources were determined under different assay methods, which makes the comparison of absolute data impossible. Therefore, it is still significant to compare the substrate tolerance of various TKs with a common real-time tracking, high-throughput assay method.

The assay methods used in specificity determination of TK enzymes are various. The conventional method for measuring TK activity uses D-ribose-5-phosphate (R5P) and D-xylulose-5-phosphate (X5P)[66] because of their superior substrate quality as acceptor and donor, respectively, and the reaction is followed spectrophotometrically by the amount of reduced nicotinamide adenine dinucleotide (NADH) consumed by auxiliary enzyme catalyzed reactions.[67, 68] Similar multi-enzyme assays have been reported using alternative acceptor or donor substrates of TK.[69-71] An umbelliferone-based sensitive fluorogenic assay has been devised to determine TK activity and stereoselectivity.[72-74] Although operated in a continuous mode like the standard method, none of these highly substrate-specific methods is adaptable to probe the kinetics of TK for other non-natural substrate analogs.

Several assay formats have been developed to probe the substrate tolerance of TK by direct measurement of remaining substrate or product formed. Measurement of chiral product formation by optical activity[75] or ellipsometry[76] is highly dependent on substrate structure and thus not of generic utility. Determination of HPA depletion by near-UV spectroscopic monitoring,[77] enzymatic assay[78] or HPLC analysis[79, 80] is hampered by low sensitivity or low throughput. Colorimetric determination of ketose formation by heterocyclic condensation reactions is limited by low specificity,[81], while tetrazolium red-based oxidation is restricted to non-hydroxylated aldehyde acceptors such as used for modified propanal tolerance;[82] in addition, both latter methods require multiple reagent

handling steps and are thus time consuming and laborious. A major disadvantage of all methods that involve substrate/product determination is, however, that they are discontinuous and do not allow direct measurement of enzyme kinetics, which not only is tedious but also a significant source of experimental error.

In the TK catalytic reaction with HPA as the donor, one equivalent of bicarbonate is produced upon HPA decarboxylation with each ketol transfer cycle by TK, which causes the pH value of the reaction system to rise (Scheme 3.1). Direct monitoring of pH changes by a pH indicator is highly convenient, and independent of the structure of an acceptor substrate. This assay principle has been successfully developed to monitor enzymatic reactions that involve various types of ester hydrolysis,[83-85] phosphoryl transfer,[86] nucleotidyl transfer[87] and glycosyltransfer[88, 89] as well as decarboxylation reactions.[90] The advantages of pH based assay methods are obvious: pH indicators are inexpensive; no auxiliary enzymes are necessary; initial reaction rates can be monitored in real-time, and the reactions should be easily adaptable to high-throughput mode in microtiter plate format. Indeed, based on this principle Hübner *et al.* have reported a pH based assay method for TK using *p*-nitrophenol as pH indicator,[91] but the low sensitivity of the latter (light yellow color development) makes this method ineffective for kinetic determinations of low enzyme activities that are typical with non-natural acceptor substrates.




Scheme 3.1. Principle of the pH based assay method for TK activity determination

In this chapter, a new continuous pH-based assay method has been developed for the high-throughput screening of TK, based on HPA as donor and phenol red as sensitive pH indicator, which is generic and can be used for the reliable colorimetric determination of specific enzyme activity and measurement of kinetic constants. This method is particularly suitable for mapping the substrate tolerance of TKs from different sources with a broad variety of natural and non-natural, α-hydroxylated and non-hydroxylated, non-phosphorylated acceptor substrates. The results obtained from kinetic measurements are verified by exemplary preparative synthesis and characterization of the corresponding chiral products. Because of its simplicity, the method should also prove valuable for the quantitative screening of mutant TK

libraries required in the directed evolution of novel enzyme properties.

## 3.2. Results and discussion

### 3.2.1. Development of pH-based assay for transketolase

Protons released or consumed in an enzymatic reaction can be detected *via* color change of a suitable pH indicator. An assay based on colorimetrical determinations of this type is a direct measure of the enzyme of interest, allowing both for rapid qualitative and, under suitably controlled conditions, also for quantitative measurements.[92] The TK catalyzed decarboxylation of HPA produces one equivalent of the corresponding bicarbonate salt per catalytic cycle. In the TK reaction described in Scheme 3.1, HPA is a medium strength acid with a p$K_a$ value lower than 2, while aldehyde acceptor and the chiral hydroxyketone produced are both neutral. At the pH optimum of the TK reaction (around 7.5)[42] none of them will affect the pH of reaction system. However, the bicarbonate ion is the dissociated form of carbonic acid, which by its effective first p$K_{a1}$ of 6.33, owing to its unfavorable dehydration equilibrium, actually is a significantly weaker acid than HPA (p$K_{a2}$ is 10.22).[93] Bicarbonate is amphoteric and involved in dynamic dissociation equilibria in water (Scheme 3.2), and its solutions react alkaline due to the partial formation of hydroxide. Therefore, the pH of the TK reaction system rises upon conversion of HPA. This pH change is revealed in the presence of a pH indicator by a color change, and the stoichiometric formation of bicarbonate can be determined photometrically for absolute quantification.


Scheme 3.2. Dynamic dissociation equilibrium of bicarbonate in water

To ensure maximum sensitivity and fidelity of a pH-dependent enzyme assay, the color change interval of the indicator needs to be around the optimal pH of the biocatalysis reaction. For our study, we chose phenol red as the pH indicator because of its high absorption coefficient ($\varepsilon$ = 56,000 $M^{-1}\cdot cm^{-1}$ at 557 nm)[86], and because its p$K_a$ of 7.4 and applicable pH range (6.8–8.2) coincide with the appropriate pH optimum of TK catalysis (around pH 7.5

depending on the source of TK).[42] Phenol red exposed to alkali changes its color from yellow to bright red and thus can be easily observed by the human eye, with sensitivity higher by at least one order of magnitude when compared to *p*-nitrophenol. The highest sensitivity results at low buffer concentrations and high indicator concentrations, but in unbuffered solution the background absorbance level changes too readily upon very small experimental variations. For our study, a buffer concentration as low as 2 mM was found to be satisfactory, and phenol red concentration at 10 mg/L (0.028 mM) was optimal for signal intensity; at least up to this concentration level no inhibition of TK activity was notable by the presence of phenol red.

For a choice of buffer, its p$K_a$ and the p$K_a$ of indicator dye should be as similar as possible to ensure that the color change is proportional to the number of produced protons.[90] TK activity is usually determined in Tris (p$K_a$ 8.06) or Gly-Gly (p$K_{a2}$ 8.25) buffer solution.[94] However, upon tests with Tris buffer we consistently noted an initial absorbance decrease of the assay solution before addition of HPA to start the reaction that continued for several minutes before reaching a steady absorption level (Figure 3.1). This decrease, which indicates a slight acidification of the solution, was only observed when the assay mixture contained Tris buffer plus any one of the aldehyde or HPA components. Apparently, the primary amine group of Tris slowly reacts with electrophilic carbonyl compounds (aldehyde and HPA) to form covalent products such as carbinolamine, imine, or oxazolidine,[95-97] which due to their lower basicity lead to a release of protons (Scheme 3.3). As Tris buffer proved unsuitable for this assay, it was switched to triethanolamine (TEA), which has a similar p$K_a$ (7.8) but lacks a reactive primary amino group. Indeed, TEA buffer produced a stable absorption signal for the assay system with no detectable pH variation from chemical reactivity and therefore was chosen for further development. Although certain tertiary amines have been shown to catalyze decomposition of HPA reminiscent to the TK reaction at elevated temperature,[98, 99] no reaction occurred with TEA under moderate conditions.

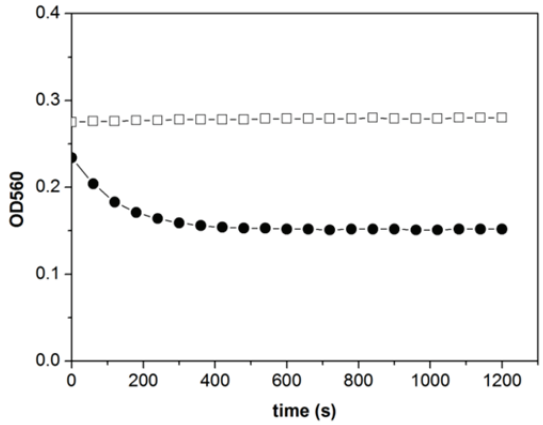
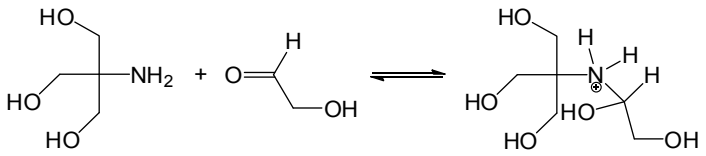



Figure 3.1. Observation of pH changes from chemical background reactivity of amine buffers. □ 2 mM TEA (pH 7.5), ● 2 mM Tris-HCl (pH 7.5). Samples contained 0.028 mM phenol red and 50 mM glycolaldehyde in 200 µL of 2 mM amine-HCl buffer at a starting pH 7.5. Absorbance was measured at 560 nm by plate reader.




Scheme 3.3. Covalent interaction between Tris and HPA

The relationship between the quantity of $HCO_3^-$ produced upon HPA conversion and the absorbance of indicator was determined by titration of the entire TK assay mixture with $NaHCO_3$, that is, with all potentially buffering elements included, but omitting an aldehyde substrate to exclude catalytic turnover. A calibration curve (Figure 3.2a) was measured at 560 nm with variation of the $HCO_3^-$ concentration from 0.0 to 0.8 mM mimicking the low degree of HPA conversion by TK under initial rate conditions. This standard curve, which shows good linearity over the entire range, was used to calculate the increase of $HCO_3^-$ concentration during the catalytic reaction. The limits of detection (LOD) and limits of quantification (LOQ) values derived from the linear regression curve[100, 101] are 0.04 mM and 0.11 mM $HCO_3^-$, respectively.

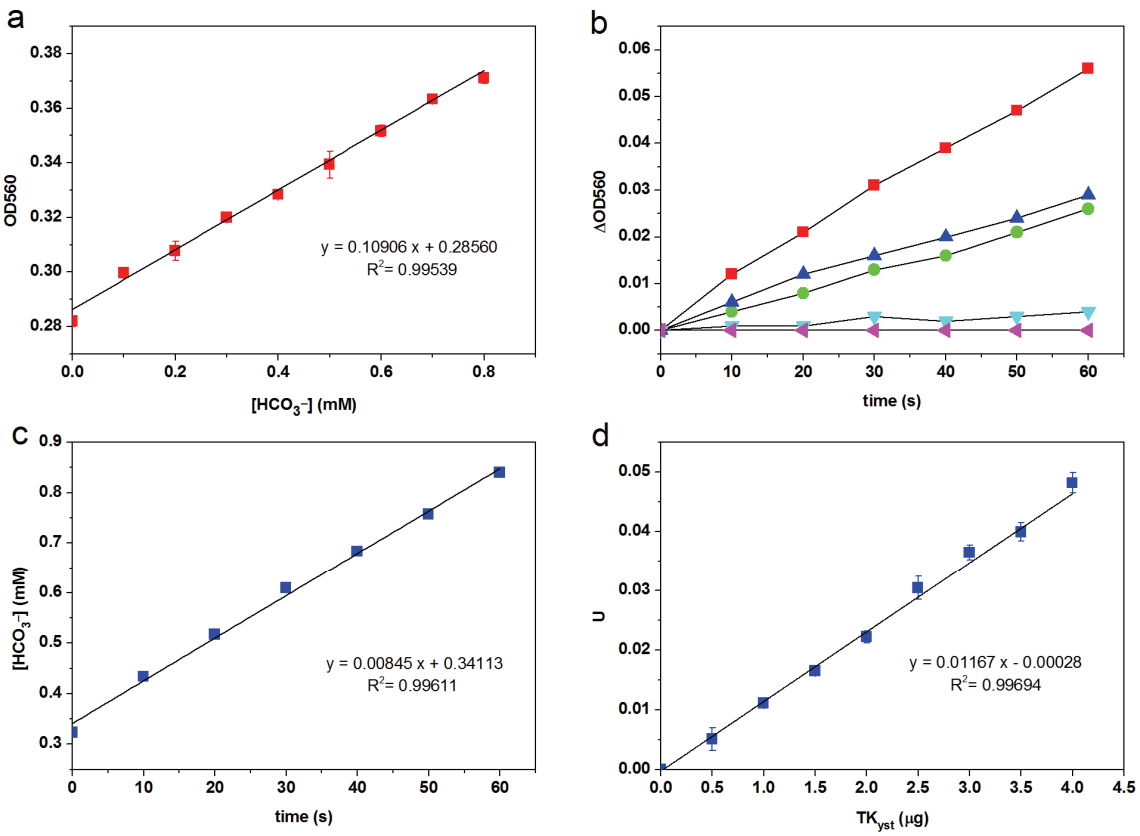



Figure 3.2. Development of the assay method. Samples contained 0.028 mM phenol red, 2.4 mM ThDP, and 9 mM $MgCl_2$ in 200 µL of 2 mM TEA buffer (pH 7.5) as well as enzyme and acceptor as specified. After addition of 50 mM LiHPA, absorbance was measured at 560 nm using a plate reader. (a) Standard curve; reaction mixtures contained 12 µg $TK_{yst}$ and varying amounts of $NaHCO_3$ (0–0.8 mM). (b) Time curve of absorbance change using different acceptors substrates. Assay solutions contained 12 µg $TK_{yst}$ and various aldehydes; ■ 50 mM D-glyceraldehyde, ● 50 mM D,L-glyceraldehyde, ▲ 50 mM D-erythrose, ▼ 200 mM D-ribose (32 µg $TK_{yst}$); ◀ control without enzyme. (c) Time curve for absolute data. Absorbance data from (b) for 50 mM D-glyceraldehyde was converted into equivalent $HCO_3^-$ concentration according to the calibration curve. The slope of the resultant curve was used for calculating the specific activity. (d) Dependence of rate on amount of enzyme; reaction mixtures contained 50 mM D-glyceraldehyde and varying amounts of 0-4 µg $TK_{yst}$.

A practical test of the assay was performed by using $TK_{yst}$ as catalyst with some non-phosphorylated hydroxyaldehydes that were reported to span a wider kinetic range of activity such as D-glyceraldehyde, D-erythrose, and D-ribose.[49, 94, 102] Resultant time curves for absorption changes (Figure 3.2b) clearly demonstrate a steady signal development and distinct activity within the acceptor series that correlates with known substrate quality determined by conventional assays.

For a quantitative interpretation, time curves of absorbance data for D-glyceraldehyde as acceptor (Figure 3.2c) were converted into absolute concentration of $HCO_3^-$ according to the calibration curve. Assuming initial rate conditions the slope of the resultant curve was used for calculating the specific activity of $TK_{yst}$ according to Formula 3.1. Thus, 1 U is defined as

the amount of enzyme required to produce 1 $\mu$mol of $HCO_3^-$ from HPA per minute under the assay conditions.

$$\text{specific activity } \mathbf{U/mg} = \frac{\text{slope} \times 10^3 \mu\mathbf{M/s} \times 60\mathbf{s} \times 200 \times 10^{-6}\mathbf{L}}{\text{amount of TK } \mathbf{mg}} \qquad \text{(Formula 3.1)}$$

Variation of the amount of $TK_{yst}$ catalyst confirmed a linear relationship between activity and amount of enzyme (Figure 3.2d), which demonstrates that this method can be used for quantitative determination of the absolute and relative specific activity of different TK enzymes. The observable LOD and LOQ values[100, 101] are 2.5 mU and 7.5 mU, respectively.

### 3.2.2. Determination of kinetic constants

In order to evaluate the data quality resulting from the pH-dependent assay format, quantitative substrate kinetics were attempted to be determined using $TK_{eco}$ and $TK_{yst}$ for a comparison against published data available from conventional assay measurements. In addition to the HPA donor, known good acceptor substrates were included such as glycolaldehyde, D-glyceraldehyde, and D-erythrose, but omitted D-ribose because of its >1 M apparent $K_M$ value.[94] The kinetic constants were determined by varying the concentration of the individual substrate (LiHPA or aldehyde) while keeping all other parameters constant. The results are listed in Table 3.1. For HPA the $K_M$ value of 5.5 mM determined with $TK_{eco}$ is almost identical to the recently published value of 5.3 mM using the same acceptor component and an HPLC method,[103] but is at variance to a value of 18 mM determined photometrically by using a phosphorylated acceptor component.[94] The $K_M$ of 3.4 mM with $TK_{yst}$ is somewhat smaller than that of 7 mM measured with a phosphorylated acceptor.[104] Regarding the acceptor components, the glycolaldehyde $K_M$ value of 13.4 mM determined with $TK_{eco}$ exactly matches the published value (13.4 mM), while those for D,L-glyceraldehyde and D-erythrose are somewhat higher (57.3 mM vs. 10 mM, and 194 mM vs. 150 mM, respectively). Taking into account that most of the published data resulted from conventional assays using phosphorylated donor and acceptor components in coupled enzyme systems, it is believed that our pH-based method yields data that are sufficiently reliable for direct comparison but easier and more rapid to determine. The results summarized in Table 3.1

demonstrate that overall catalytic efficiency ($k_{cat}/K_M$) of the wild-type TK enzymes decreases about one order of magnitude upon increasing the acceptor chain length from C2 to C4 (glycolaldehyde to D-erythrose). In particular, for glycolaldehyde $TK_{eco}$ shows significantly higher affinity and rate when compared to $TK_{yst}$, but otherwise it appears that the kinetic values of both $TK_{eco}$ and $TK_{yst}$ for D-glyceraldehyde, D,L-glyceraldehyde, and D-erythrose are rather similar.

Recently, Smith *et al*. have found that an active-site replacement D469E in $TK_{eco}$ causes an eight-fold higher activity than wild-type towards the non-hydroxylated substrate analog propanal with twofold improved stereoselectivity, while the variant H26Y leads to the enantiomeric product due to a reversal in the aldehyde attack.[105] For a broader kinetic comparison, these two enzyme variants $TK_{eco}^{D469E}$ and $TK_{eco}^{H26Y}$ have been reconstructed and purified to be included in the comparison with wild-type activity. According to our kinetic analysis the $TK_{eco}^{D469E}$ variant has a $K_M$ value (1.3 mM) for HPA that is lower than that of its parent wild type, but since $k_{cat}$ is also much reduced the overall efficiency for the donor substrate conversion remains practically unchanged. However, the variant's $K_M$ for glycolaldehyde, D-glyceraldehyde, D,L-glyceraldehyde and D-erythrose are higher than those of wild-type $TK_{eco}$ by a factor of about 2-3, and $k_{cat}$ values reduced by factors of 0.5-0.1, so that an overall reduction in catalytic efficiency results to 5-14% of the parent, which indicates that the enzyme variant has lost much of its catalytic capacity towards 2-hydroxyaldehydes as compared to its parent. Surprisingly, however, the $TK_{eco}^{H26Y}$ variant showed much lower activity than wild-type against the entire panel of substrates, quite in contrast to the reported four-fold higher activity for propanal with respect to wild-type $TK_{eco}$.[105] DNA sequencing proved the genetic construct to be correct, which excludes that undesired amino acid exchanges had been introduced at any other location in the TK protein sequence. Currently, it is unable to present a conclusive explanation as to the discrepancy between the kinetic studies here and the reported assay results concerning the same TK enzyme variant. However, it may be notable that the alleged higher activity of the $TK_{eco}^{H26Y}$ variant stands in contrast to the inability of the authors to identify this mutant protein during earlier screenings of libraries made by site-specific saturation mutagenesis.[106] Due to the low specific activity of this variant kinetic values determined for most of the substrates were found to be rather low with large experimental error and thus had to be omitted from the subsequent comparison with other TK enzymes.

Table 3.1. Determination of kinetic constants for different TK enzymes

| | $TK_{yst}$ | | | $TK_{eco}$ | | | $TK_{eco}^{D469E}$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | $K_M$ (mM) | $k_{cat}$ ($s^{-1}$) | $k_{cat}/K_M$ ($s^{-1}\cdot mM^{-1}$) | $K_M$ (mM) | $k_{cat}$ ($s^{-1}$) | $k_{cat}/K_M$ ($s^{-1}\cdot mM^{-1}$) | $K_M$ (mM) | $k_{cat}$ ($s^{-1}$) | $k_{cat}/K_M$ ($s^{-1}\cdot mM^{-1}$) |
| LiHPA | 3.4±0.3 | 28.4±0.6 | 8.27 | 5.5±0.5 | 33.2±0.8 | 6.02 | 1.3±0.2 | 9.2±0.2 | 6.85 |
| glycolaldehyde | 20.5±2.5 | 29.2±1.5 | 1.43 | 13.4±2.0 | 34.2±1.7 | 2.56 | 43.2±2.7 | 16.1±0.5 | 0.37 |
| D-glyceraldehyde | 39.9±2.4 | 24.2±0.7 | 0.61 | 47.8±3.8 | 31.8±1.3 | 0.67 | 152±16 | 7.9±0.6 | 0.05 |
| D,L-glyceraldehyde | 44.6±4.1 | 10.8±0.5 | 0.24 | 57.3±5.5 | 11.3±0.6 | 0.20 | 114±14 | 2.3±0.2 | 0.02 |
| D-erythrose | 214±17 | 39.2±2.0 | 0.18 | 194±33 | 38.4±4.0 | 0.20 | 230±35 | 2.3±0.2 | 0.01 |

### 3.2.3. Determination of acceptor specificities

Several authors have contributed to the determination of the substrate specificity of TK enzymes from different sources, and particularly the specific activities of wild type $TK_{eco}$[94] and $TK_{yst}$[104] towards various phosphorylated and unphosphorylated aldoses or aldehydes have been reported. Unfortunately, most data were obtained under very different reaction conditions using different assay methods, and therefore an absolute comparison is difficult or impossible. Thus, the new pH-based assay method has been used in an attempt to determine the catalytic capacities of $TK_{yst}$, $TK_{eco}$ and variants $TK_{eco}^{D469E}$ and $TK_{eco}^{H26Y}$ in parallel under identical reaction conditions for direct comparison towards various acceptor aldehydes.

The four different TK catalysts were tested against a larger variety of aldehydes as substrate analogs. The substrate library included aldehydes having a chain length from $C_1$ to $C_6$, with varying degree of hydroxylation. Among the 2-hydroxylated aldehydes, preference was given to those compounds bearing a (2*R*)-configuration to match the natural TK specificity. A test plate for a selected typical subset of experiments is shown in Figure 3.3, which clearly demonstrates the broad range of color intensity that developed during the reaction in individual wells, depending on the different kinds of acceptor substrates and TK catalysts. For good substrates, the bright red coloration indicates that practically complete reaction takes place in less than 30 min, while weak substrates require monitoring for longer time, e.g. up to a few hours, or use of larger enzyme quantity. The specific activity data corresponding to the quantitative evaluation for the entire substrate library is shown in Figure 3.4.

From a quantitative comparison of the specific data, it becomes apparent that the catalytic capabilities of wild-type $TK_{eco}$ and $TK_{yst}$ are quite similar with only very little variation. Both enzymes give highest rates for glycolaldehyde followed by D-glyceraldehyde, but $TK_{eco}$ shows about 20-40% higher activity, respectively, than $TK_{yst}$. Aldehydes with shorter (methanal) or longer carbon backbones (equal or larger than tetroses) cause a parallel decrease in rates with both catalysts. Common sugars (pentoses, hexoses) yielded only low specific activities even at higher substrate concentration (200 mM instead of 50 mM), which is likely due to the minimal prevalence in solution of free aldehyde forms representing the true enzyme substrate.

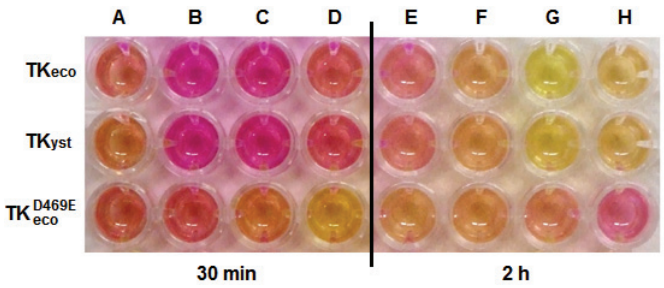



Figure 3.3. Reaction plates for TK assay. The assay solution contained 50 mM or 200 mM acceptor aldehyde, 4, 12 or 20 µg TK, 2.4 mM ThDP, 9 mM $MgCl_2$, 0.028 mM phenol red in 2 mM TEA buffer (pH 7.5). The reaction started after adding 50 mM LiHPA, total assay volume was 200 µL. Photographs were taken 30 min (left; good substrates) and 2 h (right; weak substrates) after reaction start, respectively. A. 50 mM methanal, B. 50 mM glycolaldehyde, C. 50 mM D-glyceraldehyde, D. 50 mM D,L-glyceraldehyde, E. 50 mM D,L-2,4-dihydroxybutanal, F. 200 mM D-glucose, G. 200 mM propanal, H. 200 mM ethanal.

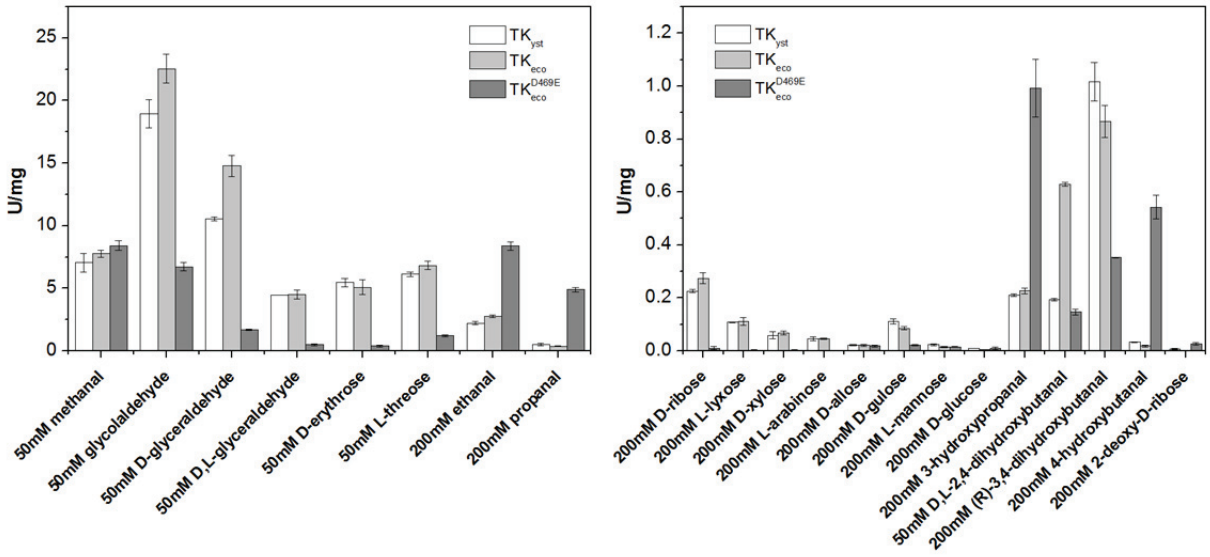



Figure 3.4. Determination of TK acceptor specificity. The assay solution contained 50 mM or 200 mM acceptor, 4, 12 or 32 µg TK, 2.4 mM ThDP, 9 mM $MgCl_2$, 0.028 mM phenol red and 2 mM TEA (pH 7.5). The reaction started after adding 50 mM LiHPA, Total assay volume was 200 µL. The OD increase was measured at 560 nm by plate reader.

Wild-type TK enzymes have a strong preference for α-hydroxylated acceptors with strict stereoselectivity towards the (2*R*)-configuration, which is clearly revealed by the results for D-glyceraldehyde versus racemic substrate. Using D,L-glyceraldehyde as an acceptor (at identical 50 mM concentration), both $TK_{eco}$ and $TK_{yst}$ show reduced activity, even smaller than 50% as compared to pure D-glyceraldehyde. Notwithstanding the lower concentration of the reactive D-enantiomer, this result might illustrate that the L-configurated isomer is acting as an inhibitor of TK. On the other hand, stereoselectivity of $TK_{eco}$ and $TK_{yst}$ towards hydroxyl configuration at C3/C4 positions is much less distinct than at C2, which is evident by the relatively similar rates for D-erythrose versus L-threose, and by the relatively small rate variation within the series of pentoses (D-ribose, L-lyxose, D-xylose and L-arabinose). Deletion of acceptor hydroxylation causes a strong drop in activity for both $TK_{eco}$ and $TK_{yst}$ enzymes; relative to glycolaldehyde the aliphatic structural analogs ethanal (2-hydroxyl deletion) and propanal (replacement of 2-OH for methyl) react more slowly by factors of 12% and 3% (even at fourfold concentration), respectively. The loss in activity upon 2-deoxygenation cannot be recovered by hydroxylation at more distant positions such as present in 3-hydroxpropanal, (*R*)-3,4-dihydroxybutanal, 4-hydroxybutanal, or 2-deoxy-D-ribose which all show rather low activity with wild-type TK enzymes.

For variant $TK_{eco}^{D469E}$, previously identified to accept propanal with eightfold higher activity than wild-type TK,[105] it consistently also shows significantly higher activities towards other 2-deoxygenated aldehydes (ethanal, propanal, 3-hydroxy-propanal, and 4-hydroxybutanal) than the corresponding wild-type enzyme. Although 2-deoxy-D-ribose was better accepted than D-ribose when compared to wild-type, interestingly the variant showed lower activity towards (*R*)-3,4-dihydroxybutanal and also for all aldehydes bearing a 2-hydroxylation, which indicates its relatively narrow substrate limitation. Unfortunately, the unexpectedly low specific activity of $TK_{eco}^{H26Y}$ rendered a direct comparison with this variant impossible.

The new assay format has allowed to rapidly recording a quantitative multi-substrate activity profile of closely related $TK_{yst}$, $TK_{eco}$, and $TK_{eco}^{D469E}$ variants. For a better visualization of their substrate scope, the relative specific activity data was processed to gray scale values according to the Reymond method[107] to draw a functional fingerprint map of the three enzyme variants for a graphical representation (Figure 3.5). The specific activity of $TK_{eco}$ towards 50 mM glycolaldehyde was used as 100% reference value. The resulting map indeed readily illustrates the similarities and differences in the enzymes' relative substrate specificities. Thus, it clearly indicates the high catalytic conformity of the wild-type TK enzymes, while variant $TK_{eco}^{D469E}$ acts rather differently. This feature is remarkable as $TK_{yst}$ shows only 42.2% identity to $TK_{eco}$ in its protein sequence, whereas variant $TK_{eco}^{D469E}$ is 99.8% identical to $TK_{eco}$, except for the

single amino acid substitution. The fingerprint also clearly reveals that beyond the functional similarity $TK_{eco}$ displays somewhat higher activities towards glycolaldehyde and D-glyceraldehyde than $TK_{yst}$, whereas $TK_{eco}^{D469E}$ should be a preferred catalyst towards methanal and, in particular, 2-deoxygenated aldehydes (ethanal, propanal, 3-hydroxpropanal, and 4-hydroxybutanal).

Table 3.2. Determination of TK acceptor specificity

| Acceptor | Conc. (mM) | $TK_{yst}$ (U/mg) | $TK_{eco}$ (U/mg) | $TK_{eco}^{D469E}$ (U/mg) |
|---|---|---|---|---|
| methanal | 50 | 7.06±0.75 | 7.77±0.28 | 8.41±0.37 |
| glycolaldehyde | 50 | 19.0±1.1 | 22.6±1.2 | 6.72±0.32 |
| ethanal | 200 | 2.23±0.11 | 2.79±0.11 | 8.38±0.32 |
| D-glyceraldehyde | 50 | 10.6±0.2 | 14.8±0.8 | 1.69±0.05 |
| D,L-glyceraldehyde | 50 | 4.47±0.00 | 4.50±0.35 | 0.497±0.080 |
| 3-hydroxypropanal | 200 | 0.211±0.005 | 0.227±0.011 | 0.993±0.109 |
| propanal | 200 | 0.503±0.093 | 0.370±0.020 | 4.90±0.16 |
| D-erythrose | 50 | 5.47±0.33 | 5.08±0.58 | 0.393±0.080 |
| L-threose | 50 | 6.12±0.18 | 6.84±0.32 | 1.21±0.06 |
| D,L-2,4-dihydroxybutanal | 50 | 0.193±0.005 | 0.631±0.007 | 0.147±0.011 |
| (*R*)-3,4-dihydroxybutanal | 200 | 1.02±0.07 | 0.867±0.060 | 0.353±0.001 |
| 4-hydroxybutanal | 200 | 0.034±0.001 | 0.019±0.004 | 0.543±0.045 |
| D-ribose | 200 | 0.227±0.007 | 0.275±0.020 | 0.011±0.006 |
| L-lyxose | 200 | 0.109±0.001 | 0.111±0.014 | n.d.* |
| D-xylose | 200 | 0.060±0.013 | 0.068±0.008 | n.d. |
| L-arabinose | 200 | 0.045±0.007 | 0.046±0.002 | n.d. |
| 2-deoxy-D-ribose | 200 | 0.007±0.002 | n.d. | 0.027±0.004 |
| D-allose | 200 | 0.022±0.003 | 0.021±0.003 | 0.019±0.005 |
| D-gulose | 200 | 0.112±0.010 | 0.086±0.006 | 0.022±0.003 |
| L-mannose | 200 | 0.023±0.003 | 0.015±0.002 | 0.015±0.001 |
| D-glucose | 200 | 0.009±0.001 | 0.005±0.001 | 0.011±0.006 |

*n.d. = not detectable

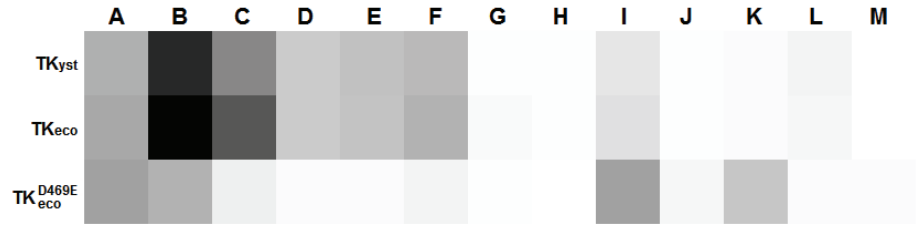



Figure 3.5. Fingerprint comparison for $TK_{yst}$, $TK_{eco}$, and $TK_{eco}^{D469E}$. The relative specific activity data was converted to gray scale value.[107] The specific activity of TKeco towards 50 mM glycolaldehyde was used as 100% reference. A. 50 mM methanal, B. 50 mM glycolaldehyde, C. 50 mM D-glyceraldehyde, D. 50 mM D,L-glyceraldehyde, E. 50 mM D-erythrose, F. 50 mM L-threose, G. 50 mM D,L-2,4-dihydroxybutanal, H. 200 mM D-ribose, I. 200 mM ethanal, J. 200 mM 3-hydroxypropanal, K. 200 mM propanal, L. 200 mM (*R*)-3,4-dihydroxybutanal, M. 200 mM 4-hydroxybutanal.

Although the number of substrates included in the fingerprint analysis (considering factors such as chain length, hydroxylation pattern, and sense of chirality) and the number of enzymes available in this study are rather limited, it appears that this approach should prove a valuable tool to discriminate TK enzyme variants that might become available from various other sources or from experiments in directed evolution. While in this first attempt it has been only considered the influence of chirality inherent to the substrate, it is noted that the level of chiral induction at C3 in the product may further be indicated by color-coding for *(R)* or *(S)* configuration[107] once this is determined.

## 3.3. Conclusion

Using HPA as the donor component in TK reactions causes a defined pH increase, the rate of which has been shown to be directly proportional to the activity of TK. The proton consumption can be revealed in the presence of the pH indicator phenol red accurately with high specificity and sensitivity, which allows using this system for quantitative measurements covering a broad range of TK activity. TEA buffer at low 2 mM concentration proved sufficient to avoid an undesired pH interference of assay components.

It has been demonstrated that this new assay format is applicable to the reliable determination of kinetic constants using three different kinds of TK, including wild-type and mutated protein variants. The method is also useful for a rapid quantitative multi-substrate screening to chart and distinguish the distinct acceptor specificities of enzyme variants, resulting in a functional fingerprint map of their substrate tolerance that readily illustrates the

similarities and differences in the enzymes' relative specificities. This pH based assay should prove as a very useful tool in the screening of novel TK enzymes for their substrate tolerance and for their modification by directed evolution because of its simplicity, speed, high sensitivity, and low costs.

# 4. Directed Evolution of Thermostable Transketolase from *Geobacillus stearothermophilus*

## 4.1. Introduction

The TK from *Geobacillus stearothermophilus* ($TK_{gst}$) is a novel thermostable enzyme which has been cloned and identified recently within a French collaboration.[65] This thermophilic TK shows impressive high activity and stability at elevated temperatures. It has revealed that it has an optimum temperature range around 60–70°C and retains almost 100% activity for one week at 50°C and for 3 days at 65°C, which allows carrying out an easy, one-step purification by simple heat shock treatment of crude cell extracts at 70°C. When comparing its performance under different temperatures, the reaction rate of $TK_{gst}$ with glycolaldehyde showed an 8–14 fold increase at 70°C compared to 20°C, and a four-fold higher activity at 50°C when compared to $TK_{eco}$ under the same conditions. When tested at 50°C with other aldehyde acceptors, the $TK_{gst}$ activity was approximately 3 times higher than that observed at 20°C.[65]

High thermostability is a significant and useful feature for an enzyme to expand its potential for preparative applications and future biocatalyst development. Processes at elevated temperatures could offer opportunities to extended TK biocatalysis applications, improving substrate solubility of hydrophobic aldehyde acceptors and enzyme tolerance towards non-conventional substrates. A number of applications of this new $TK_{gst}$ in biocatalysis have been performed with Li-HPA as donor and glycolaldehyde, D-glyceraldehyde and butyraldehyde as acceptors, from which the corresponding products L-erythrulose, D-xylulose and 1,3-dihydroxyhexan-2-one were obtained, respectively, leading to identical chiral induction (3*S*-configuration) of products for α-hydroxylated acceptors and improved *ee* values for products from non-hydroxylated substrates.[65]

As a key enzyme involved in carboligation, TK has been proved as an efficient biocatalyst for carbon chain extension and introduction of a chiral hydroxylation site. Upon conversion of non-hydroxylated substrates, the TK catalytic reaction can form deoxyketose-type products, which are regarded as valuable intermediates for the pharmaceutical industry.[108] However,

as discussed in Chapter 3, in comparison with the α-hydroxylated substrates, non-α-hydroxylated aldehydes cannot be well accepted by wild-type TK. TK's specific activity towards non-α-hydroxylated substrates is usually about 40 folds lower than for the corresponding α-hydroxylated derivatives.[108] Probably, this is because remoral of the α-hydroxy group reduces the hydrogen bonding contacts between TK and acceptors, leading to higher $K_M$ values, more freedom of substrate motion and lower activation of substrate in the TK's active pocket, and consequently lower catalytic efficiency of TK. Thus, an improved structural adjustment at active site by mutagenesis may help to improve binding and activation of non-α-hydroxylated substrates. Recently, Hailes *et al.* have constructed single site saturation mutagenesis libraries on 10 structurally defined position in the active site and 10 phylogenetic ones for $TK_{eco}$ to improve its acceptance towards non-phosphorylated substrates[58] and in particular for the non-α-hydroxylated propanal.[59] Among these 20 mutation positions, D469 showed the highest plasticity in comparison with other ones. A change of Asp469 to Thr, Tyr, Ala and Glu gave 4.3 to 8 fold activity improvement towards propanal.[59] It should be mentioned that mutagenesis also modified — occasionally enhanced or even reversed — $TK_{eco}$'s stereoselectivity, especially providing mutant D469E with selectivity for 90% (*S*)-configuration and variant H26Y for 88% (*R*)-configuration.[61]

$TK_{gst}$ has 50.2% identity with $TK_{eco}$ and almost an identical active site constitution. Because of its high thermostability and better stereoselective performance than $TK_{eco}$, it seems plausible that $TK_{gst}$ and its potential descendents could become ideal biocatalysts for the synthesis of deoxyketose-type products. In this chapter, the structure of $TK_{gst}$ has been simulated and aligned with those of $TK_{eco}$, $TK_{yst}$ and $TK_{ban}$ by computational methods to analyze the contributions for substrate binding and catalysis. According to this structure alignment, two critical amino acid residues were then chosen for building single site and double site saturation mutagenesis libraries in the quest for improved $TK_{gst}$ activity towards non-α-hydroxylated acceptors, as evaluated by using a pH-based high throughput screening method.

```
            1                                                                80
TKeco     (1) ----------------MSSRKELANAIRALSMDAVQKAKSGHPGAPMGMADIAEVLWRDFLKHNPQNPSWADRDRFVLSN
TKyst     (1) ------------MAQFSDIDKLAVSTLRLLSVDQVESAQSGHPGAPLGLAPVAHVIFKQLRCN-PNNEHWINRDRFVLSN
TKban     (1) MYKTFTNISEGKNSMSHSIEQLSINTIRTLSIDAIEKANSGHPGMPMGAAPMAYTLWTQFMKHNPNNPTWFNRDRFVLSA
TKgst     (1) --------------MAHSIEELAITTIRTLSIDAIEKAKSGHPGMPMGAAPMAYTLWTKFMNHNPANPNWFNRDRFVLSA

            81                                                              160
TKeco    (65) GHGSMLIYSLLHLTGYDLPMEELKNFRQLHSKTPGHPESGVTPLGVETTTGPLGQGIANAVGMAIAEKTLAAQFNRPGHD
TKyst    (68) GHSCALLYSMLHLLGYDYSIEDLRQFRQVNSRTPGHPEF-HSAG-VEITSGPLGQGISNAVGMAIAQANFAATYNEDGFP
TKban    (81) GHGSMLLYSLLHLSGYDVTMDDLKNFRQWGSKTPGHPEYGHTAG-VDATTGPLGQGIATAVGMAMAERHLAAKYNRDAYN
TKgst    (67) GHGSMLLYSLLHLSGYDVSMDDLKQFRQWGSKTPGHPEYGHTPG-VEATTGPLGQGIAMAVGMAMAERHLAATYNRDGFE

            161                                                             240
TKeco   (145) IVDHYTYAFMGDGCMMEGISHEVCSLAGTLKLGKLIAFYDDNGISIDGHVEGWFTDDTAMRFEAYGWHVIRDIDG-HDAA
TKyst   (146) ISDSYTFAIVGDGCLQEGVSSETSSLAGHLQLGNLITFYDSNSISIDGKTSYSFDEDVLKRYEAYGWEVMEVDKGDDDME
TKban   (160) IVDHYTYAICGDGDLMEGVSAEASSLAAHLQLGRLVVLYDSNDISLDGDLNRSFSESVEDRYKAYGWQVIRVEDG-NDIE
TKgst   (146) IINHYTYAICGDGDLMEGVASEAASLAGHLKLGRLIVLYDSNDISLDGELNLSFSENVAQRFQAYGWQYLRVEDG-NNIE
```


Sequence 4.1. Protein sequence alignment of TK from *E. coli*, *S. cerevisiae*, *B. anthracis* and *G. stearothermophilus*. Yellow: identical; blue: conservative; green: weakly similar.

## 4.2. Results and discussion

### 4.2.1. Sequence and structure analysis of $TK_{gst}$

$TK_{gst}$ has a highly similar protein sequence to TKs from *E. coli*, *S. cerevisiae* and *B. anthracis*, for which protein crystal structures have been determined.[51, 52, 109-112] Among them, $TK_{gst}$ and $TK_{ban}$ have the highest similarity to each other due to the close genetic relationship between *Geobacillus* sp. and *Bacillus* sp.. Thus, the crystal structure of $TK_{ban}$ was chosen as the template for the simulation of the 3D structure of $TK_{gst}$. The computational simulation was created by using the SWISS-Model web tool,[113, 114] which also recommended $TK_{ban}$ as an optimum template from alignment of protein sequence data and a PDB database search. The simulated 3D structure is shown in Figure 4.1a, from which one can appreciate its high similarity to that of $TK_{ban}$, and even $TK_{eco}$ and $TK_{yst}$. Focusing on their active sites, all these four TKs share the same 'boot' shape active pocket structure (Figure 4.1b and 4.2). The "feet

space" in this "boot" is for the binding of the cofactor ThDP, which is surrounded by 11 amino acid residues (Ile249, His68, Gly116, Pro117, Leu118, Tyr441, Glu162, Gly158, Asp157, Gly156 and Asn187). The "bootleg" is constituted by nine amino acid residues (His462, Arg520, Asp470, His28, Gly264, His263, Leu192, Leu382 and Ser385), forming a channel for substrate entrance to the catalytic core. Therefore, the modification of the amino acid residue components forming the "bootleg" may change the TK's acceptance towards acceptors.

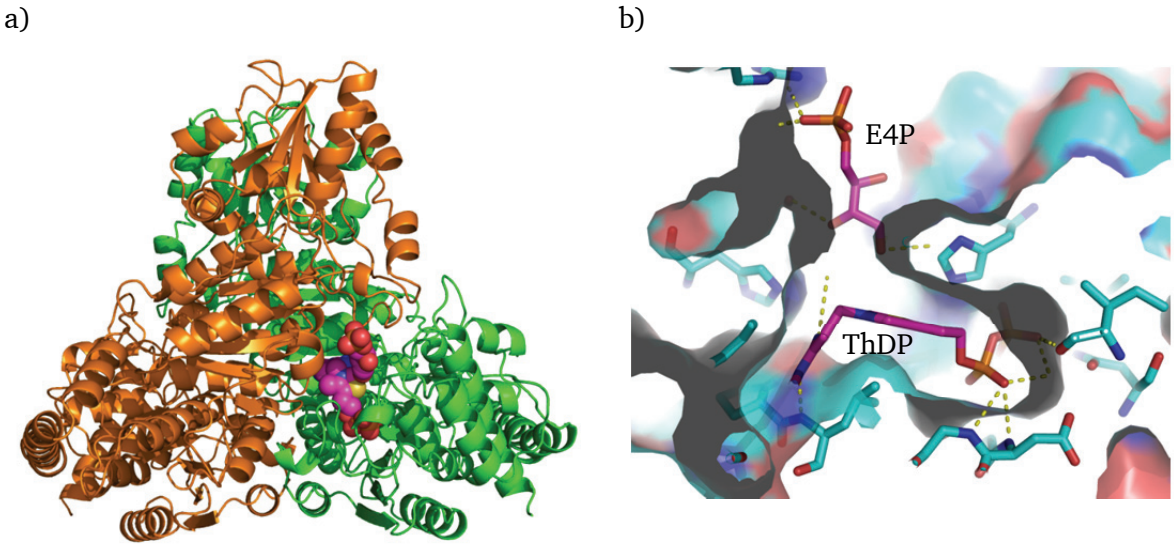



Figure 4.1. Simulated structure of $TK_{gst}$ dimer. a) Overall structure of $TK_{gst}$. b) The "boot" shaped active pocket of $TK_{gst}$. E4P: D-erythrose 4-phosphate; ThDP: thiamine pyrophosphate.

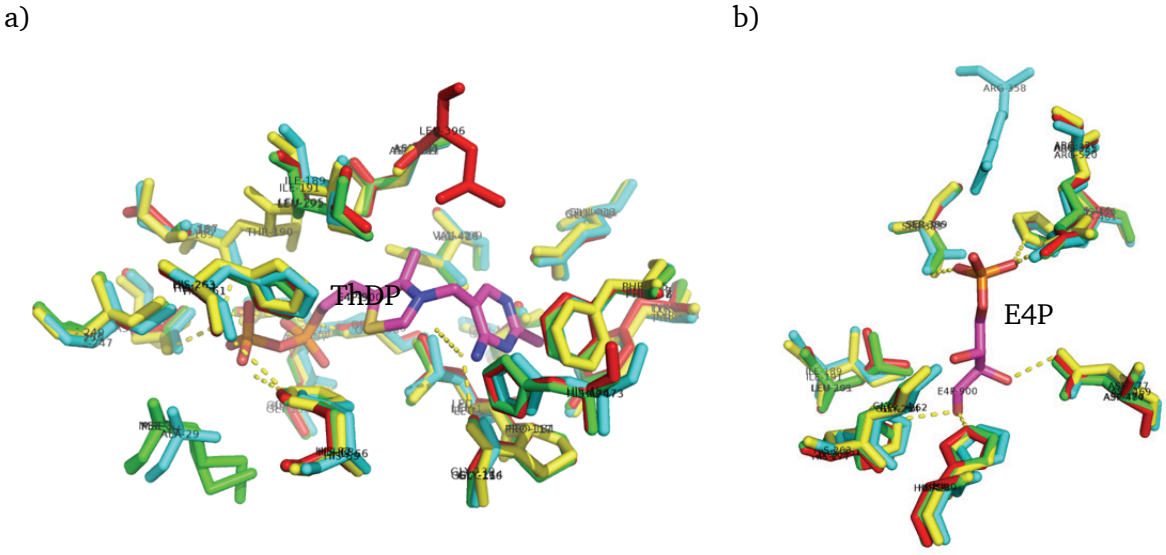



Figure 4.2. Active site alignment of $TK_{gst}$ with those of $TK_{eco}$, $TK_{yst}$ and $TK_{ban}$. a) ThDP binding pocket. b) E4P binding pocket. Red: $TK_{yst}$; Blue: $TK_{eco}$; Yellow: $TK_{ban}$; Green: $TK_{gst}$. E4P: D-erythrose 4-phosphate; ThDP: thiamine pyrophosphate.

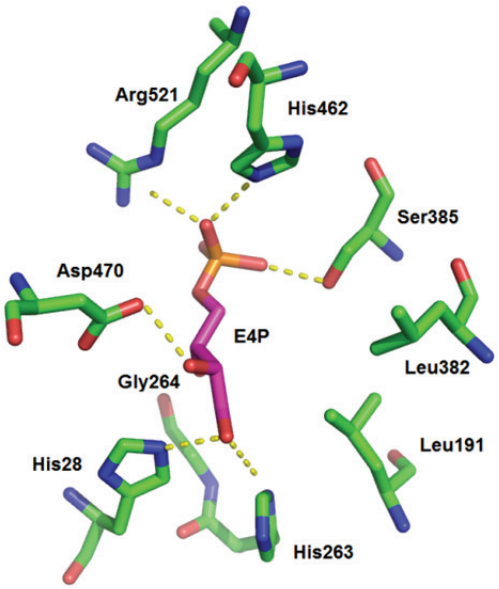



Figure 4.3. Model of acceptor binding pocket of $TK_{gst}$, using E4P substrate conformation determined in $TK_{yst}$. E4P: D-erythrose 4-phosphate. Residue numbering for $TK_{gst}$.

The earlier saturation mutagenesis study of $TK_{eco}$ towards non-α-hydroxylated acceptors had been performed by Hailes *et al.* with propanal as a typical substrate and using a colorimetric screening method.[59, 61, 115] This study was focused on both the ThDP binding and acceptor pockets. As a result of the screening, several positive mutants have been found on Asp469, Ala29, Asp259 and His26 positions, among which D469E was the top variant with an eight fold increased activity towards propanal as compared to the wild-type.[59, 61] D469 is the amino acid residue that directly contributes a hydrogen bond to the α-hydroxy group of an α-hydroxylated acceptor. The replacement of Asp469 by other amino acid residues with longer side chains such as Glu may narrow the acceptor access tunal, and further limit the mobility of bound substrates, and thus improve the catalytic efficiency with non-α-hydroxylated acceptors such as propanal (Figure 4.3). $TK_{gst}$ also has the equivalent Asp at position 470. Thus, it is considered that saturation mutagenesis on this position could produce similar results to those obtained by Hailes *et al.*. Therefore, Asp470 was chosen as one mutantion site.

On the other hand, Leu382 is located at a position almost opposite to Asp470 (Figure 4.3). Although it seems to have no specific contribution to substrate binding and activation in the wild-type structure, modification of this position may complement an effect of Asp470 mutagenesis. Furthermore, variation at the Leu382 position may also provide an opportunity for reversing TK's enantioselectivity, for example to accept L-glyceraldehyde (which rescently has been ifentified to be an inhibitor to TKs[108]) and forming the products containing the

non-natural, reversed configuration. According to the above structure analysis, Asp470 and Leu382 were selected as the most plausible mutation positions for single and coupled site mutagenesis library construction.

### 4.2.2. Building of $TK_{gst}$ mutagenesis libraries

Saturation mutagenesis is an efficient mutagenesis strategy based on the structure and interaction analysis between enzyme and substrate, which has been successfully used to improve the properties of an enzyme, including stability, substrate tolerance and stereoselectivity.[26] This method was also chosen here to construct three libraries to induce mutations on Leu382 or/and Asp470 sites, including individual two single-site libraries on each site and a double-site library corering both positions with NNS codon strategies.

For the single-site libraries at each Asp470 and Leu382, the gene codons at the mutation site were designed as NNS in order to reduce the library size to 95 colonies, while maintaining a 95% coverage of all 20 amino acids,[26] which would readily be compatible with a 96-well plate formate. For the coupled Asp470 and Leu382 double site library, there are two well-known construction strategies. One is named ISM (Iterative Saturation Mutagenesis),[116] in which the most active mutants are identified at the Asp470 or Leu382 positions in a first round of screening, followed by raudomizing the second site and using the first positive mutants as templates. This is a relative simple but efficient method, which has the advantage of small library size, especially when a few mutant sites need to be induced. However, unfortunately potentially interesting mutants might be missed with this method, because one or more mutant sites are fixed at an early stage by certain residues in these libraries. This will limit the diversity of mutants, because it ignores that the enzyme property could be improved by a cooperative interaction between any two or more amino acid residues. The other strategy performs a saturation mutagenesis on both positions simultaneously using an NNS codon (Table 4.1) on both Leu382 and Asp470 sites, at least 3,066 colonies must be included in this library to achieve a statistical 95% coverage with 20 residues.[26] Although the total size of this library is much larger than that of a single site one, it is still feasible to handle the library on 31 96-well plates. Therefore, NNS degeneracy was also chosen for the construction of the double-site library.

**Table 4.1.** Codon degeneracy of saturation mutagenesis.[26]

| Codon degeneracy | No. of codons | No. of AA | No. of stops | AA encoded | 95% coverage for 2 pos. | 95% coverage for 3 pos. |
|---|---|---|---|---|---|---|
| NNK/NNS | 32 | 20 | 1 | All 20 | 3066 | 98163 |
| DBK | 18 | 12 | 0 | ARCGILMFSTWV | 969 | 17470 |
| NDK | 12 | 12 | 0 | RNDCGHILFSYV | 430 | 5175 |
| NRT | 8 | 8 | 0 | RNDCGHSY | 190 | 1532 |

For the construction of these three libraries the commercial QuikChange method was used, which is an ideal one-step mutant induction strategy that does not require any ligation process (Figure 4.4).[117, 118] Two pairs of primers were designed including an NNS codon for single site libraries. The PCR was found to be a difficult step, in which the annealing temperature had to be optimized up to 68 or 72 °C instead of recommended 60 °C to achieve suffricient PCR amplification. In the standard QuikChange method, the *Dpn*I treated PCR product is used directly to transform the expression host cells, followed by the screening on selection plates. Finally, the positive colonies are picked as library members, from which a few recombinants are chosen for sequencing to evaluate the degree of randomization of the mutations.[117, 118] However, the generated PCR products are self-annealing plasmids only with two remaining nicks because of the absence of ligase. Therefore, their transformation efficiency is much lower than that of the corresponding continuous ones, which may create an insufficient number of positive colonies on selection plates. On the other hand, choosing only a few random samples for evaluation of the diversity of a mutagenesis library may not factually reflect the statistical degree of true mutation randomness for the whole library. In addition, sequencing multiple samples is quite costly.

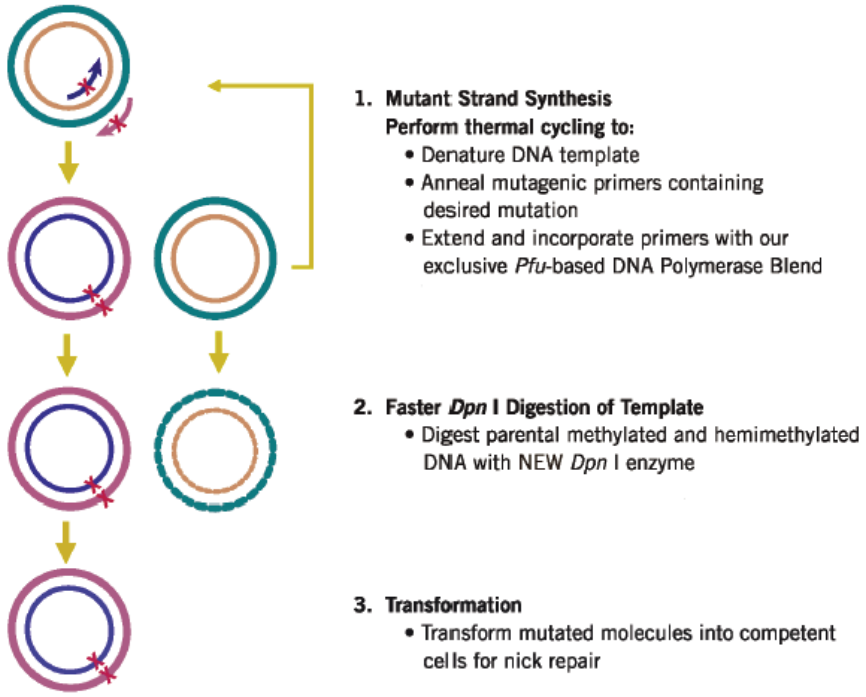
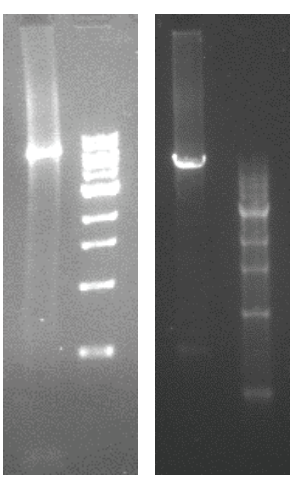


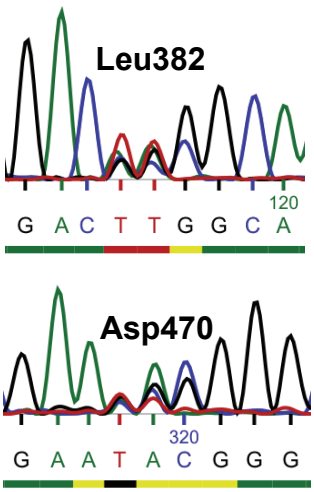

Figure 4.4. Principle of the QuikChange mutagenesis method. (Copied from the manual of the QickChange mutagenesis kit from Agilent, USA)

In order to deal with these problems, the process after *Dpn*I digestion was modified. The PCR product was transformed into competent XL1 cells followed by liquid cultivation in selection LB medium instead of spreading on plates. The mixture of cells was harvested after overnight cultivation and their internal plasmids were extracted as a mixture sample. Owing to the same transformation rate and change of each mutant, this mixed plasmid sample contained all the variants with the same proportion as produced during the mutagenic PCR. Thus, this sample was sent for sequencing to determine the base composition on each position. Figure 4.5 shows the obtained sequencing results. The original codon TTG (Leu382) and GAC (Asp470) have been successfully replaced by NNS, in which N was shown as A/T/G/C overlapping peaks and S as G/C overlapping peaks. Although the A/T/G/C peaks cannot be quantitatively compared between each position, their peak heights or peak areas have a quantitative relationship for the specific site. The highest peak indicates that its represented nucleotide residue dominates this position. Zheng *et al* also used this method to evaluate their improved mutagenesis method.[119] By calculating the heights of overlapping peaks, A/T/G/C were found to be present with a similar coverage of around 20-25% for N and G/C 50% for S, which are close to the ideal value of 25% and 50%. This result proved that the nucleotide residue randomness at the two mutated positions was excellent for library building. After this evaluation, the

plasmid mixture was transformed into competent BL21(DE3) cells. Because the plasmids had been repaired during the plasmid amplification, the efficiency of the second transformation was definitively sufficient for the single site library building on a single plate. Finally, 96 colonies for each single site library were picked.

a) b) c)






| Leu382 | 1 | | 2 | | 3 | |
|---|---|---|---|---|---|---|
| | Height % | Area% | Height % | Area% | Height % | Area% |
| A | 24 | 25 | 26 | 30 | 3 | 1 |
| C | 20 | 19 | 22 | 15 | 33 | 37 |
| G | 17 | 16 | 23 | 23 | 63 | 62 |
| T | 39 | 40 | 30 | 32 | 1 | 0 |

| Asp470 | 1 | | 2 | | 3 | |
|---|---|---|---|---|---|---|
| | Height % | Area% | Height % | Area% | Height % | Area% |
| A | 22 | 18 | 42 | 41 | 1 | 0 |
| C | 19 | 20 | 20 | 19 | 56 | 58 |
| G | 28 | 27 | 25 | 26 | 38 | 36 |
| T | 31 | 35 | 13 | 14 | 6 | 5 |

Figure 4.5. Evaluation of single site libraries. a) PCR product identification by agar gel. Left for Lib-Leu382; right for Lib-Asp470. b) Sequencing results of the mixed plasmid samples. The original codon TTG (Leu382) and GAC (Asp470) have been successfully replaced by NNS, in which N was shown as A/T/G/C overlapping peaks and S as G/C overlapping peaks. c) Calculation of peak heights or peak areas from the sequencing results.

The double site library construction was attempted in two ways. The first used the QuikChange Multi-mutagenesis Kit with a modified QuikChange method that uses two single primers for each mutation site and a thermophile ligase for the cyclization of single nucleotide chain. The second method performs independent single site mutagenesis sequentially to achieve the double site modification (Figure 4.6a). The former is easy to be operated and also timesaving. However, because the PCR products from each amplification round cannot be used as templates for the second round, very little PCR products can be obtained as compared to a normal PCR. Although the PCR conditions were optimized many times, probably because of the latter reason, PCR products could not be achieved in sufficient quantities for further steps. Thus, the second method was used to perform single site mutagenesis consecutively, in which the evaluated plasmid mixture from the first round of mutation was used as the template for the second round of PCR (Figure 4.6b). The evaluation of these two rounds of mutagenesis by DNA sequencing is presented in Figure 4.7. The nucleotide randomness at each mutation site was also excellent for library building. Finally, 3,456 colonies (36 plates) were chosen for the double site library.

a) b)

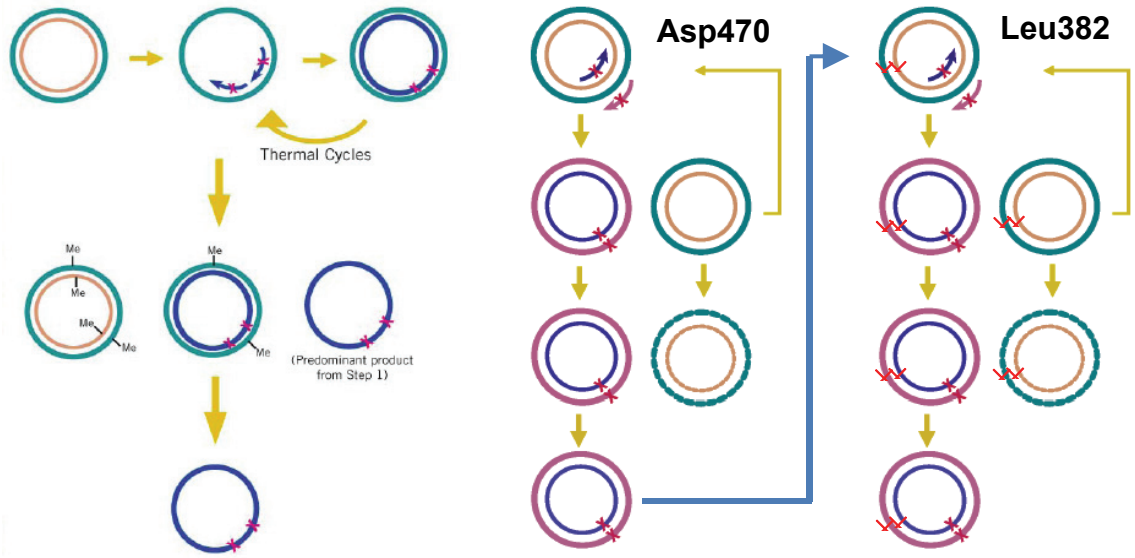
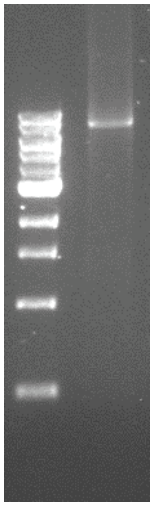
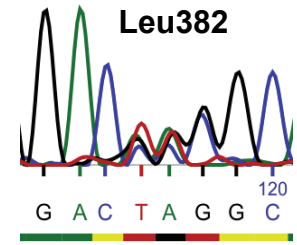
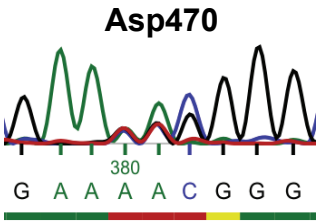



Figure 4.6. Principle of double site library construction. (Modified based on the copy from the manual of the QickChange mutagenesis kit from Agilent, USA)

a) b) c)






| Leu382 | 1 | | 2 | | 3 | |
|---|---|---|---|---|---|---|
| | Height % | Area% | Height % | Area% | Height % | Area% |
| A | 26 | 26 | 30 | 32 | 0 | 0 |
| C | 16 | 12 | 16 | 15 | 43 | 43 |
| G | 22 | 25 | 27 | 27 | 50 | 52 |
| T | 36 | 37 | 27 | 26 | 7 | 5 |

| Asp470 | 1 | | 2 | | 3 | |
|---|---|---|---|---|---|---|
| | Height % | Area% | Height % | Area% | Height % | Area% |
| A | 25 | 21 | 40 | 42 | 2 | 0 |
| C | 22 | 22 | 19 | 18 | 58 | 63 |
| G | 26 | 28 | 20 | 19 | 35 | 30 |
| T | 26 | 29 | 20 | 21 | 6 | 6 |

Figure 4.7. Evaluation of double site library. a) PCR product identification by agar gel. b) Sequencing results of the mixed plasmid samples. The original codon TTG (Leu382) and GAC (Asp470) have been successfully replaced by NNS, in which N was shown as A/T/G/C overlapping peaks and S as G/C overlapping peaks. c) Calculation of peak heights or peak areas from the sequencing results.

### 4.2.3. Development of a high-throughput screening method for TK mutagenesis library

Tetrazolium-based colorimetric[82] and HPLC assays[80] have been applied for the screening of $TK_{eco}$ mutagenesis libraries. However, the former can only be used for non-α-hydroxylated aldehyde acceptors, and both methods are rather laborious and time-consuming. Thus, the pH-based assay was further evaluated for its suitability for the high-throughput screening of TK libraries. Assays for HTS not only require small sample volume, high throughput, and robustness, but also adequate sensitivity, reproducibility, and accuracy in order to discriminate among a large number of substrates that span the entire range of activity.

The culture and lysis methods were optimized for recombinant cells on 96-well plates, and adapted to the pH-based assay for the screening in microtiter plate format. For proof-of-principle, only recombinant cells expressing wild-type $TK_{yst}$ and $TK_{eco}$ were used as samples, for which BL21(DE3) host cells served as "negative" control; background activity from the native TK gene of the host was deemed too low to interfere with the measurements. In view of the requirement for very low buffer concentration in the assay system, cell lysis had to be effected in the absence of high concentration of buffering components. For this purpose, a recently published method was adopted by using 1/10 BugBuster solution with inclusion of 0.5 mg/mL lysozyme and 4 U/mL Benzonase endonuclease as the resuspension buffer.[89] By shaking at room temperature for 0.5-1 h, practically all *E. coli* host cells became disintegrated as judged by microscopic examination. A 40 $\mu$L sample of the supernatant was applied for activity measurement. Low concentrations of glycolaldehyde (5 mM for $TK_{yst}$, 2.5 mM for $TK_{eco}$) were chosen as a rate-limiting factor.

For an evaluation and validation of high-throughput screening assays, Zhang *et al.* defined a screening window coefficient called *Z*-factor,[120] a dimensionless statistical parameter that is significant for both the assay signal dynamic range and the data variation associated with the signal measurements, which therefore is suitable for assay quality assessment. The assay method is judged to be excellent when $1.0 > Z \geq 0.5$ (an ideal method is characterized by a *Z*-factor of 1). To evaluate our screening method, 48 parallel samples (one half of a 96-well plate) were used for each cell type to compare the differences within the same sample dataset as well as between that of the diverse recombinants, and calculated the *Z*-factor to assess the high-throughput application of our screening method as listed in Table 4.2. Soluble expression of $TK_{yst}$ in *E. coli* is compromised by formation of inclusion bodies and thus the

enzyme reaches a much lower soluble concentration level than attainable for $TK_{eco}$, which is why its activity towards 5 mM glycolaldehyde (2.3 mU) in crude extracts already reaches the LOD of this method (2.5 mU). However, the *Z*-factor value (0.52) is still above 0.5, which demonstrates that the pH-based assay has enough precision for its use towards a high-throughput purpose. In comparison, the application of the assay to the screening of $TK_{eco}$ variants at even lower substrate concentration is far superior.

**Table 4.2.** Evaluation of pH-based screening method

| | $TK_{yst}$ | $TK_{eco}$ |
|---|---|---|
| glycolaldehyde (mM) | 5.0 | 2.5 |
| average activity (mU) | 2.3 | 38.3 |
| relative standard deviation | 14% | 7% |
| *Z*-factor | 0.52 | 0.80 |

### 4.2.4. Screening, identification and docking of positive $TK_{gst}$ variants towards propanal

#### 4.2.4.1. Screening of positive $TK_{gst}$ variants towards propanal

The screening of $TK_{gst}$ mutagenesis libraries was carried out on 96-well plates with an optimized pH-based high throughput assay method, in which 40 $\mu$l lysate sample was used, and with the assistance of a multichannel pipetting work station, Biomek 2000 from Beckman (USA). The increasing $OD_{560}$ during the initial 10 min was recorded by a Vmax plate reader and further converted into the absolute reaction rate to compare the activity discrepancy between each well. About 15% (192) top positive clones were picked into new plates for a second round of rescreening. Their plasmids were extracted for sequencing. Although the double site library includes the two single site libraries, the screening of two single site libraries was deemed appropriate to help in evaluating the assay method firstly because of their small size, and also to illuminate the particular effects of the individual single site mutation. The top 40 mutants from this second round of screening with propanal were finally collected for purification and independent activity identification.

### 4.2.4.2. Purification and activity identification of positive $TK_{gst}$ variants towards propanal, butanal and methoxyethanal, including their thermostability determination

The recombinant $TK_{gst}$ mutants are His-tag labeled, so they can be easily purified by Ni-affinity chromatography. However, the simultaneous purification of 40 samples by affinity chromatography is not convenient to operate. The ultrafiltration for removing high imidazole concentrations for the pH assay is a time-consuming step. Fortunately, the $TK_{gst}$ has a high thermostability, and indeed a one-step heating method could be simply performed to obtain high purity $TK_{gst}$ samples.[65] Although any change of its protein sequence may enhance its flexibility under high temperatures and thus destabilize its structure, its high thermostability is expected to remain largely during the mutagenesis of $TK_{gst}$. Therefore, the one-step heating method was also carried out for the purification of these 40 variants. Their protein concentration and purity were determined by the Bradford method and SDS-PAGE, respectively, followed by an activity determination towards propanal, butanal, methoxyethanal, and glycolaldehyde (Table 4.3).

Table 4.3. Screening results for mutant libraries of $TK_{gst}$*

| mutant position | | | | propanal | butanal | methoxyethanal | glycolaldehyde | thermo-stability |
|---|---|---|---|---|---|---|---|---|
| Leu382 | | Asp470 | | | | | | |
| AA | codon | AA | codon | 200 mg, U/mg | 200 mM, U/mg | 200 mM, U/mg | 10 mM, U/mg | $T_m$, °C |
| Leu | TTG | Ile | ATC | 1.3 | 1.4 | 1.9 | 0.65 | 74 |
| Leu | CTC | Leu | CTC | 0.97 | 1.0 | 1.6 | 0.37 | 74 |
| Phe | TTC | Asp | GAC | 0.91 | 0.86 | 1.0 | 0.70 | 72 |
| Leu | TTG | Thr | ACG | 0.71 | 0.72 | 1.2 | 0.64 | 73 |
| Asn | AAC | Phe | TTC | 0.70 | 0.72 | 0.93 | 0.20 | 71 |
| Phe | TTC | Thr | ACG | 0.65 | 0.75 | 1.0 | 0.45 | 73 |
| Leu | TTG | Tyr | TAC | 0.61 | 0.67 | 0.74 | 0.19 | 75 |
| Glu | GAG | Ser | TCG | 0.58 | 0.63 | 0.66 | 0.17 | 72 |
| Leu | TTG | Glu | GAG | 0.55 | 0.54 | 0.79 | 0.24 | 73 |
| Ile | ATC | Val | GTC | 0.54 | 0.52 | 1.1 | 0.46 | 74 |
| Val | GTC | Leu | CTG | 0.53 | 0.52 | 0.70 | 0.35 | 74 |
| Asn | AAC | Trp | TGG | 0.44 | 0.52 | 0.65 | 0.25 | 70 |
| Leu | TTG | Ser | TCG | 0.44 | 0.43 | 0.56 | 0.24 | 72 |
| Asn | AAC | Val | GTC | 0.43 | 0.42 | 0.62 | 0.38 | 71 |
| Ala | GCG | Ile | ATC | 0.39 | 0.40 | 0.57 | 0.28 | 70 |
| Phe | TTC | Ser | TCC | 0.36 | 0.38 | 0.40 | 0.15 | 70 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Phe | TCC | Asn | AAC | 0.34 | 0.39 | 0.58 | 0.48 | 71 |
| Phe | TTC | Pro | CCG | 0.33 | 0.36 | 0.45 | 0.21 | 71 |
| Leu | TTG | Glu | GAG | 0.31 | 0.30 | 0.33 | 0.09 | 71 |
| Leu | TTG | Trp | TGG | 0.31 | 0.30 | 0.38 | 0.14 | 69 |
| Ile | ATC | Leu | CTG | 0.28 | 0.30 | 0.31 | 0.07 | 70 |
| Ile | ATC | Glu | GAG | 0.28 | 0.25 | 0.28 | 0.25 | 71 |
| Asn | AAC | Ser | TCC | 0.27 | 0.30 | 0.37 | 0.10 | 71 |
| Met | ATG | Val | GTC | 0.27 | 0.28 | 0.46 | 0.10 | 69 |
| Gln | CAG | Ile | ATC | 0.26 | 0.30 | 0.45 | 0.14 | 69 |
| Asn | AAC | Ser | TCC | 0.23 | 0.23 | 0.20 | 0.06 | 69 |
| Ala | GCC | Val | GTG | 0.23 | 0.25 | 0.41 | 0.13 | 69 |
| Met | ATG | Leu | TTG | 0.23 | 0.24 | 0.35 | 0.13 | 69 |
| Phe | TTC | Val | GTG | 0.20 | 0.22 | 0.45 | 0.29 | 68 |
| Ile | ATC | Ser | TCC | 0.18 | 0.21 | 0.28 | 0.03 | 70 |
| Gly | GGG | Gly | GGC | 0.17 | 0.21 | 0.22 | 0.03 | 71 |
| Ser | TCG | Thr | ACC | 0.16 | 0.16 | 0.21 | 0.04 | 71 |
| Thr | ACG | Leu | CTC | 0.16 | 0.15 | 0.18 | 0.04 | 69 |
| Gln | CAG | Ser | TCC | 0.16 | 0.18 | 0.24 | 0.12 | 69 |
| Leu | TTG | Thr | ACG | 0.15 | 0.18 | 0.32 | 0.10 | 69 |
| Ser | TCC | Ser | AGC | 0.14 | 0.16 | 0.17 | 0.04 | 71 |
| Gly | GGG | Asp | GAC | 0.09 | 0.10 | 0.16 | 0.25 | 72 |
| **Leu**** | TTG | Asp | GAC | 0.08 | 0.11 | 0.22 | 0.41 | 71 |

* Same amino acids are marked with same color.
** The last line is for the wild-type $TK_{gst}$.

From the activity determination of all 40 variants towards propanal, it became apparent that the activity variation corresponding to mutation in Asp470 position was realized by similar exchanges as those that had been found in $TK_{eco}$. The exchanges D470E, D470T and D470W were also found in the top list of $TK_{gst}$. Remarkably, however, besides them novel variants D470I, D470L and D470S were identified in the screening. All of them presented an up to 17 fold higher activity than the wild-type. Mutation of the Leu382 position was found to be much more conservative and gave much less contribution to the activity improvement as compared to that of the Asp470 site. Only L382F showed a roughly 10 fold activity improvement. The double site library could offer almost all amino acid residue combinations at both positions of Leu382 and Asp470. The most notable combination is that when both sites contain non-polar residues, for example D470I, D470L, L382I/D470V, L382V/D470L and L382A/D470L (Figure 4.8). In this case, the 382 and 470 residues can jointly form a hydrophobic pocket for binding the alkyl moiety of propanal (Figure 4.9). When either position is a polar residue, usually one of these two positions contains a residue with a large side group, for example Phe, Thr, Tyr and Pro, which probably can narrow the tunnel of the active site entrance and limit the

mobility of propanal during the carboligation reaction. Interestingly, no basic amino acid residues were found at either of these two positions in any positive variant.

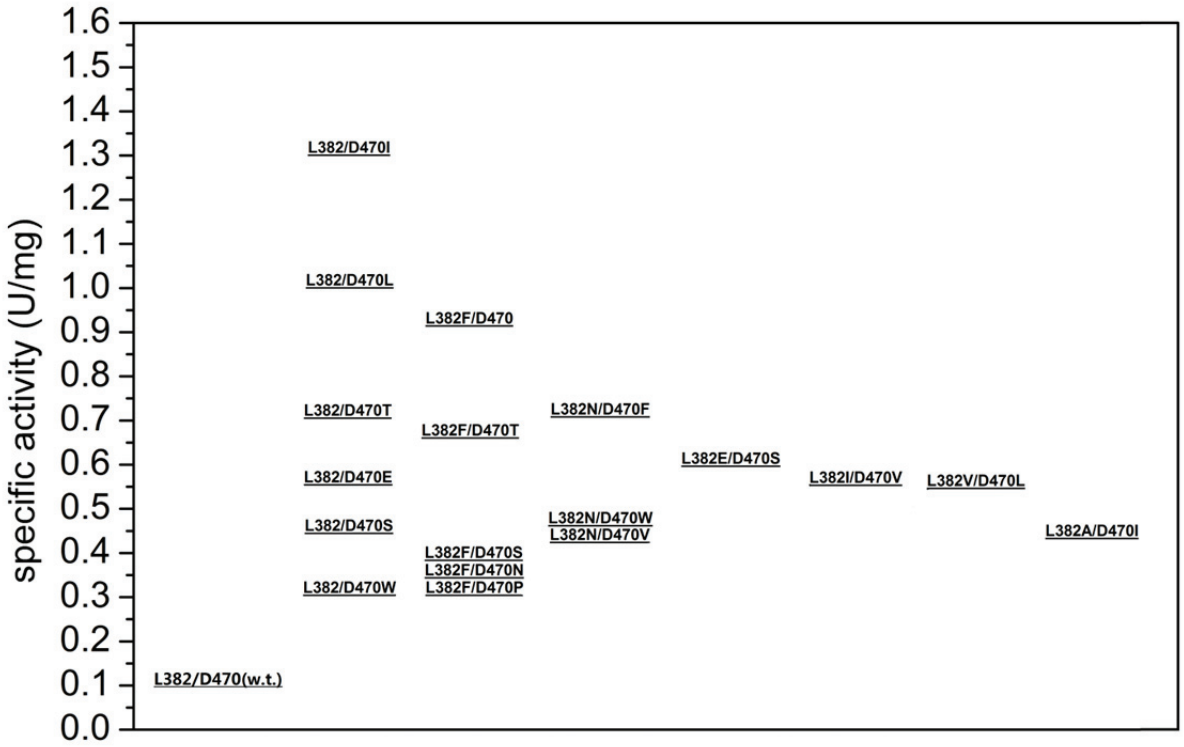



Figure 4.8. Positive variants from the screening of the $TK_{gst}$ libraries against simple aliphatic aldehydes.

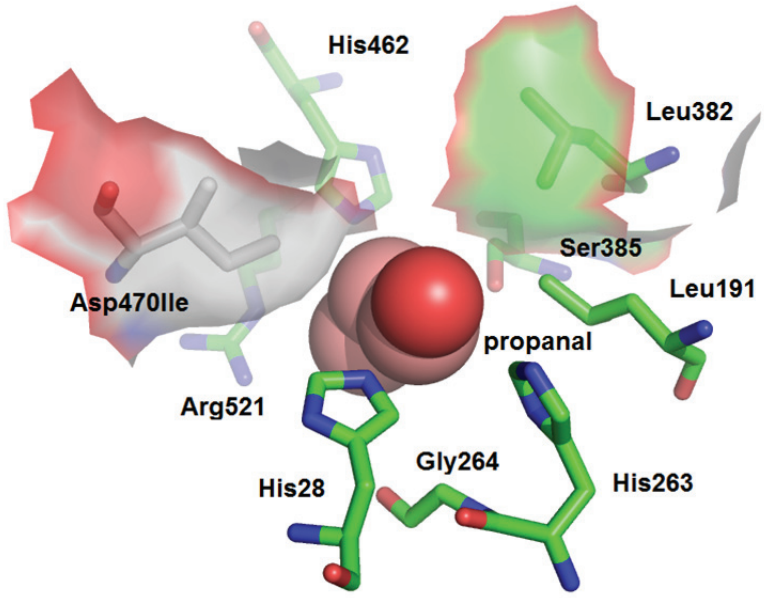



Figure 4.9. Docking of propanal as substrate analog for $TK_{gst}^{D469I}$. Asp470Ile and Leu382 form a non-polar surface to bind the aliphatic chain of propanal with higher affinity.

The specific activity of the variants towards the related aliphatic substrates butanal and methoxyethanal was also determined. The top mutants showed similar activity towards butanal in comparison with that of propanal. The methoxy group has an electronwithdrawing effect, which enhaces the electrophilicity of the carbonyl group and makes it more reactive against a nucleophilic attack. Therefore, the activity of the variants towards methoxyethanal is universally higher than that obserred with propanal and butanal. Surprisingly, the top four variants even improved or retained their activity towards glycolaldehyde. That means that the α-hydoxy group of glycoladehyde might have a different binding position which does not affect the catalytic efficiency of TK.

#### 4.2.4.3. Thermostability determination of positive $TK_{gst}$ variants

The positive variants found in the screening of $TK_{gst}$ libraries were expected to have a somewhat modified thermostability relative to the wild-type. To evaluate the effect of the individual mutagenesis events on the overall protein stability, the thermostability of all 40 variants were determined by a thermofluor-based high-throughput method,[121, 122] in which the hydrophobic fluoroprobe SYPRO orange is used as dye. During the denaturation of the protein at high temperature, its hydrophobic interior will become exposed to the probe in an aqueous solution, leading to a sharp decrease in quenching of the dye, so that a readily detected fluorescence emission can be studied in a RT-PCR experiment as a function of temperature.[122] This thermofluor-based assay showed that interestingly all of the positive variants indeed had an increased $T_m$ compared to the wild-type, and none of them had significantly lost the high thermostability. The melting temperature of the top hit reached 74 °C.

## 4.3. Conclusion

In this chapter, the saturation mutagenesis libraries of the thermostable $TK_{gst}$ at Leu382 and Asp470 have been constituted in a first attempt to improve its activity towards non-α-hydroxylated aldehydes. Leu382 and Asp470 are the corresponding residues for binding the hydroxymethylene group of α-hydroxylated aldehydes. The mutation of these two positions contained several positive variants from both single-site libraries and the double-site library.

D469I was identified as the top mutant with 17, 13 and 8.6 fold activity improvement towards propanal, butanal and methoxyethanal, respectively, in comparison to the wild-type, but still keeps high thermostability with a $T_m$ of 74 °C.

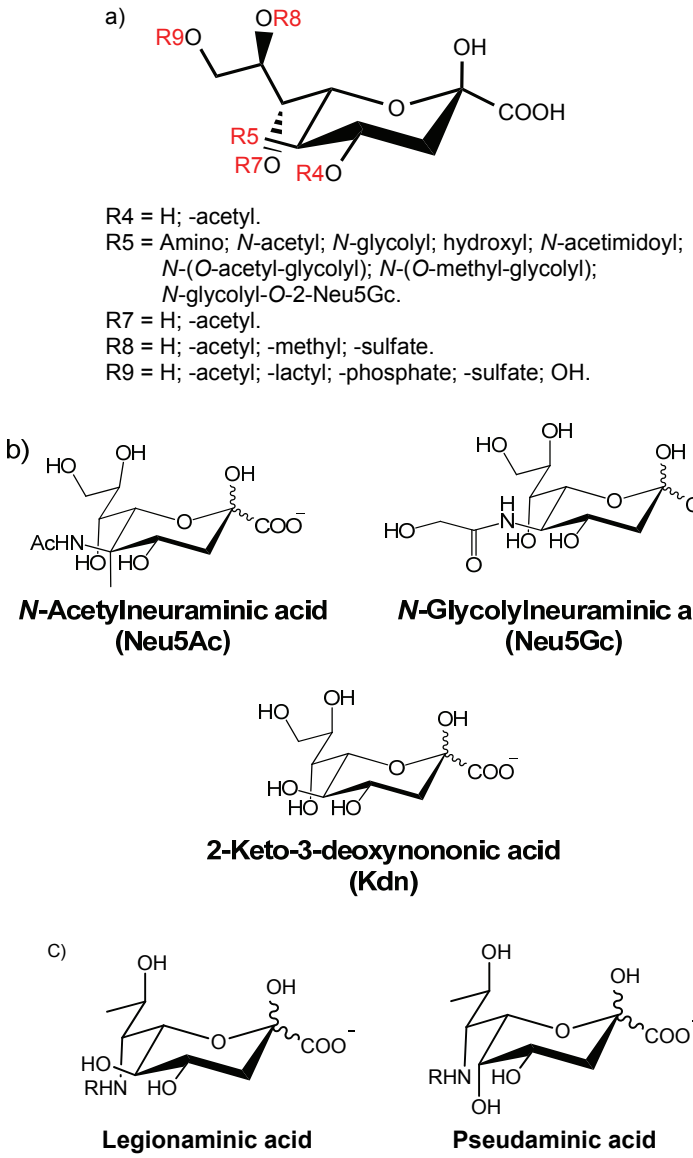
# Part II. Chemoenzymatic Synthesis of neo-Sialoconjugates

# 5. Introduction for Part II

Carbohydrates not only serve as intermediates in storing and generating energy, but also as signaling effectors, recognition markers, and structural components *in vivo*.[35] They are abundant on the surface of cells in the form of glycoproteins, glycolipids or free glycans, involved in the interaction between cells, and cells and proteins. Therefore, the biosynthesis and functions of glycoconjugates are studied with increasing attention because of their growing role in medicine and biotechnology.[35] In particular, sialic acids and sialoconjugates are a type of saccharides that play a key role in glycobiology.

## 5.1. Sialic acids and sialoconjugates

Sialic acids are a family of mostly 9-carbon sugars containing more than 50 natural members, which occur broadly in eucaryotes and prokaryotes, as well as in some viruses.[123] The parent sialic acid is neuraminic acid (Neu),[124] in which C4, C5, C7, C8 and C9 positions can be substituted or modified by hydroxyl, acetyl, amino, glycolyl, acetimidoyl, or even phosphate and sulfate groups (Scheme 5.1a).[35, 123] Therefore, the structures of sialic acids show quite high diversity. Among them, *N*-acetylneuraminic acid (Neu5Ac), *N*-glycolylneuraminic acid (Neu5Gc) and deaminoneuraminic acid (Kdn) are the most abundant sialic acid in higher organisms (Scheme 5.1b).[124] Besides, some related nonulosonic acids, such as legionaminic acid and pseudaminic acid, are also common in bacteria.[124] (Scheme 5.1c)




Scheme 5.1. Typical structures of sialic acid. a) Natural occurrence of modified sialic acid.[35] b) Structures of most common sialic acids: Neu5Ac, Neu5Gc and Kdn. C) Structures of bacterial sialic acids: legionaminic acid and pseudaminic acid.

Sialic acids are mainly present at the terminus of the free or buoconjugates glycan of glycosphingolipids and glycoproteins on the surface of cells, where they are involved in very important biological functions by their interation with different Sia-recognizing proteins in cellular adhesion and recognition, blood components and clotting, immunology and infection, transmission of nerve impulses, hormonal regulation, and many more.[123] Bacteria, especially pathogenic bacteria, also produce sialylated lipooligosaccharides or capsular polysaccharides to make against the immune response pathway of their host.[124] Therefore, both eukaryotes and prokaryotes process their own metabolism pathway to sialic acids.

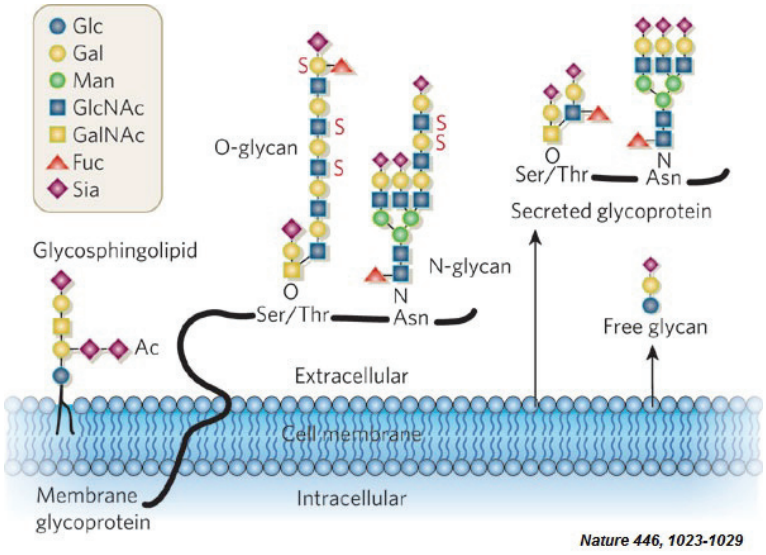
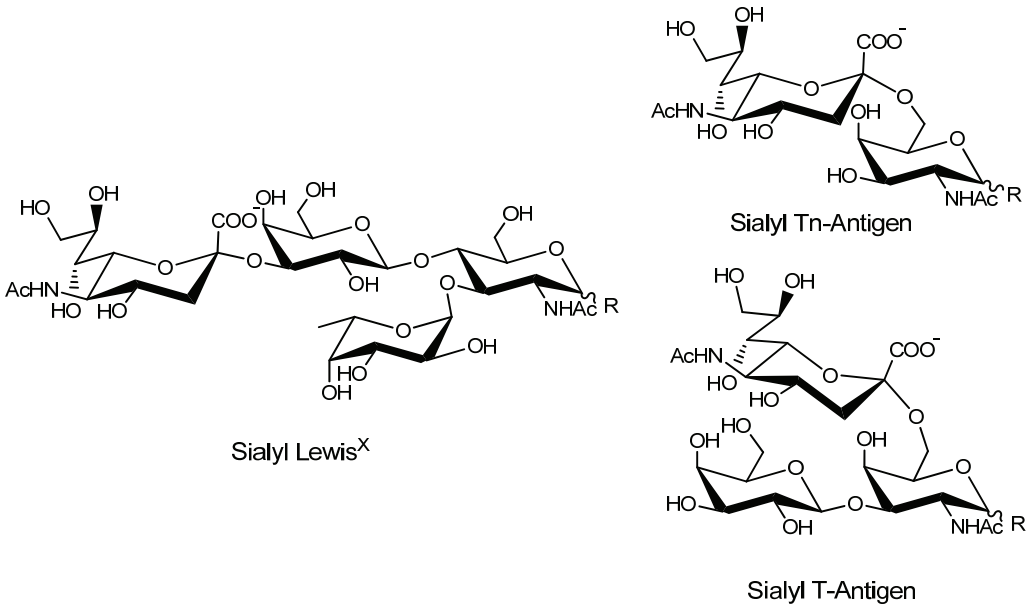



Figure 5.1. Structure model of glycans on the cell surface. (Copied from ref. [125])

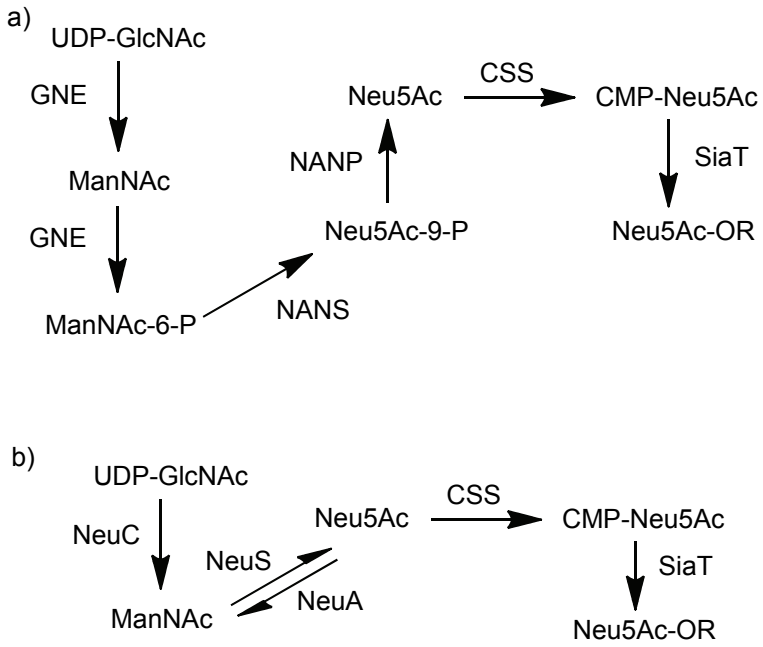
The diversity of natural sialic acid linkages and sialylated glycans is also abundant. The sialyl-associated monosaccharides are commonly galactose (Gal), *N*-acetylgalactosamine (GalNAc), *N*-acetylglucosamine (GlcNAc) and also sialic acid.[123] Sialic acids can be transferred to the Gal moiety with α-2,3-, 2,4- and 2,6-linkage, to a GalNAc moiety with α-2,6-linkage, to a GlcNAc moiety with α-2,4- and 2,6-linkage, and to a sialic acid moiety with α-2,8- and 2,9-linkage, even forming polysialic acids.[123] Notably, Sialyl Lewis$^X$ (SL$^X$, Sia$\alpha$2-3Gal$\beta$1-4(Fuc$\alpha$1-3)GlcNAc$\beta$OR) is an important sialylated epitope involved in inflammation and cancer metastasis.[123] The structure change of sialylated glycans is related to many physiological responses. For example, the shorter *O*-linked GalNAc and βGal-(1-3)-αGalNAc epitope is often aberrantly highly produced on the surface of cancer cells and are capped by sialic acid residues, while can be utilized for tumor typing in diagnostics.[126-128] Owing to these important roles of sialic acid in the cellular activity, the systematic study of the structure of sialosides and corresponding cellular responses have a significant value for the understanding of the mechanism of carbohydrate-protein interactions in physiological events for diagnostic and therapeutic targets. Therefore, it is of high importance to be able to synthesize uniform sialoconjugates *in vitro* for research on sialic acid in glycobiology and for medicinal purposes.




Scheme 5.2. Structures of Sialyl Lewis$^X$ and Sialyl T/$T_n$ antigen.

## 5.2. Biosynthetic metabolism of sialic acid and sialosides *in vivo*

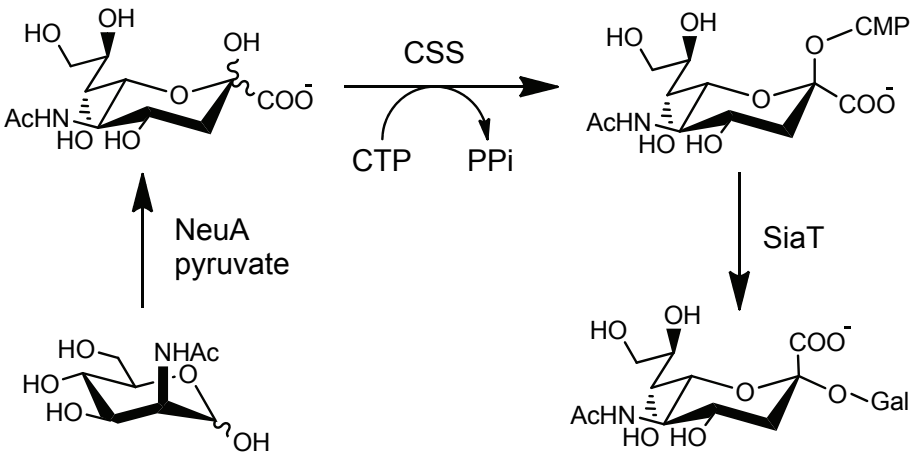
Although sialic acids are partly shared by eukaryotes and prokaryotes, the biosynthesis of sialic acid in eukaryotes and bacteria is quite different because of the difference of cell structure and metabolism. In eukaryotes, the biosynthesis of sialic acid starts from UDP-GlcNAc. A bifunctional hydrolyzing UDP-GlcNAc 2-epimerase/ManNAc-6-kinase (GNE) converts UDP-GlcNAc to ManNAc, followed by the phosphorylation of ManNAc to give ManNAc-6-P. ManNAc-6-P is then converted to Neu5Ac-9-phosphate by Neu5Ac-9-phosphate synthase (NANS) and dephosphorylated by Neu5Ac-9-phosphate phosphatase (NANP).[129] Correspondingly, in some bacterial pathogens, UDP-GlcNAc is converted to ManNAc by hydrolyzing UDP-GlcNAc 2-epimerase (NeuC). However, in the following step, ManNAc is used directly by Neu5Ac synthase (NeuS) to form Neu5Ac in the presence of phosphoenolpyruvate (PEP). Moreover, some bacteria can also scavenge sialic acid from host sialosides or utilize free sialic acids directly from their host.[129] In both eukaryotes and bacteria, the CMP-activation of sialic acid is catalyzed by CMP-Neu5Ac synthetase (CSS). The CMP-activated sialic acid can be accepted by sialyltransferase (SiaT), which transfers the sialic acid moiety to the Gal, GalNAc, GlcNAc or sialic acid terminus of glycans.




Scheme 5.3. Biosynthetic pathway of sialic acid (a) in eukaryotes and (b) in prokaryotes. GNE: hydrolyzing UDP-GlcNAc 2-epimerase/ManNAc-6-kinase, NANS: Neu5Ac-9-P synthetase, NANP: Neu5Ac-9-P phosphatase, CSS: CMP-sialic acid synthetase, SiaT: sialyltransferase, NeuC hydrolyzing UDP-GlcNAc 2-epimerase, NeuS: sialic acid synthase, NeuA: sialic acid aldolase.

## 5.3. Chemoenzymatic synthetic route of sialic acid and sialoconjugates *in vitro*

The synthetic pathway of sialosides *in vitro* begins with Man(NAc) or its chemically synthesized derivatives. In the presence of pyruvate, sialic acid aldolase catalyzes the reversible condensation of Man(NAc) and pyruvate producing the corresponding sialic acids. Because the chemical modification of Man(NAc) can be relative easily induced, a multitude of sialic acid derivatives can be obtained by this approach, whereby the natural sialic acids and their derivatives carrying with modifications at the C5, C7, C8 and C9 positions have been successfully synthesized.[130-137]




Scheme 5.4. Enzymatic synthesis of sialosides *in vitro.* NeuA: sialic acid aldolase, CSS: CMP-sialic acid synthetase, SiaT: sialyltransferase.

Due to the difficulty of chemical sialylation,[124] enzymatical methods are mostly applied for the CMP-activation and sialylaction of glycans, respectively. The activation step is catalyzed by CMP-sialic acid synthetase (CSS, E.C. 2.7.7.43), which uses CTP as nucleotide donor and $Mg^{2+}$ as cofactor.[138] Besides animals, CSS can also be found in many microorganisms, especially in the pathogenic ones, such as *Clostridium thermocellum*, *Streptococcus agalactiae*, *Escherichia coli*, *Haemophilus ducreyi*, *Mannheimia haemolytica, Pasteurella haemolytica, Neisseria meningitides*,[139] and newly found CSS from *Pasteurella multocida*.[140] The CSS from *Neisseria meningitidis* ($CSS_{nme}$) is a typical enzyme which has been cloned[141], crystalized for structure determination and investigated for synthetic applications.[142] Its kinetic constants have been well determined by a pH-based assay method and interpreted by the acceptor binding alignment with CSS from *Mus musculus* ($CSS_{mmu}$).[143] $CSS_{nme}$ has a relative broad acceptor tolerance towards 5', 7', 8' and 9' modified sialic acids.[141, 143] A series of Neu5Ac analogues with the modification on the acylamino group could be accepted by $CSS_{nme}$, though their corresponding $k_{cat}/K_M$ values were much lower than that of Neu5Ac.[143]

Sialyltransferase (SiaT) catalyzes the transfer of CMP-activated sialic acid to a suitable terminal glycan moiety. Eukaryote species normally have multiple SiaTs. However, mammalian enzymes normally cannot be well expressed in prokaryotic expression systems, which limits their applications for the synthesis *in vitro*. Bacterial SiaT are often found in pathogenic bacteria, and most enzymes can be expressed in soluble form in an *E. coli*

expression system. Thus, these have been developed as efficient tools for sialoside synthesis. According to their target glycosidic bond types, bacterial SiaTs are mainly classified into α/β-galactoside α-2,3-sialyltransferase (2,3SiaT, E.C. 2.4.99.4) and β-galactoside α-2,6-sialyltransferase (2,6SiaT, E.C. 2.4.99.1). Up to the present, several bacterial SiaTs have been cloned from *Campylobacter jejuni*, *Pasteurella multocida*, *Haemophilus* sp., *Neisseria* sp., *Photobacterium* sp., and *Vibrio* sp., which are classified in the CAZy families 42, 52 and 80, respectively.[129] Some of them also seem to have sialidase and trans-sialidase activites.[144]

The substrate specificities of some bacterial SiaTs have been determined either by employing radioactive [145] or fluorescent labeled CMP-Neu5Ac [146], or by direct analysis for formed product by HPLC.[147] Bacterial 2,3SiaT can accept both α- and β-galactosides, but apparently the 2,6SiaT can only transfer to β-galactosides. From the crystal structure determination of marine bacterial SiaTs, the active sites of $2,3SiaT_{pph}$ and $2,6SiaT_{psp}$ have been identified, as well as their catalytic mechanism. [148-150] The synthesis of sialosides can be carried out as a one-pot multi-enzyme process, as demonstrated by a series of acceptors including galactosides, lactosides, and LacNAc, as well as $Le^x$, T- and $T_n$-antigens, and oligosaccharide chains.[129]

Table 5.1. Bacterial sialyltransferases characterized so far.

| SiaT | resource | activity | CAZy family | ref. |
|---|---|---|---|---|
| CstI | *Campylobacter jejuni* | 2,3SiaT | 42 | [151] |
| CstII | *Campylobacter jejuni* | 2,3/8SiaT | 42 | [151] |
| CstIII | *Campylobacter jejuni* | 2,3SiaT | 42 | [151] |
| Lic3B | *Haemophilus influenzae* | 2,3/8SiaT | 42 | [152] |
| LsgB | *Haemophilus influenzae* | 2,3SiaT | 52 | [153] |
| HD2,3ST | *Haemophilus ducreyi* | 2,3SiaT | 80 | [154] |
| Lst | *Neisseria meningtitidis* | 2,3/6SiaT | 52 | [155] |
| Lst | *Neisseria gonorrhoeae* | 2,3/6SiaT | 52 | [155] |
| PmST1 | *Pasteurella multocida* | 2,3/6SiaT | 80 | [147] |
| PmST2 | *Pasteurella multocida* | 2,3SiaT | 52 | [156] |
| PmST3 | *Pasteurella multocida* | 2,3SiaT | 42 | [157] |
| Bst | *Photobacterium damselae* | 2,6SiaT | 80 | [145, 158] |

| | | | | |
|---|---|---|---|---|
| Plst6 | *Photobacterium leiognathi* JT-SHIZ-119 | 2,6SiaT | 80 | [159] |
| Plst6 | *Photobacterium leiognathi* JT-SHIZ-145 | 2,6SiaT | 80 | [160] |
| PpST | *Photobacterium phosphoreum* | 2,3SiaT | 80 | [161] |
| Pst3-224 | *Photobacterium* sp. JT-ISH-224 | 2,3SiaT | 80 | [162] |
| Pst6-224 | *Photobacterium* sp. JT-ISH-224 | 2,6SiaT | 80 | [162] |
| 2,3ST | *Vibrio sp. JT-FAJ-16* | 2,3SiaT | 80 | [163] |

## 5.4. Sialoside-based carbohydrate microarray and dendrimers

Due to the complex structure and multi-biofunction of oligosaccharides, the glycan and its acceptors become important targets for understanding many cell physiologies (e.g. cell adhesion, recognition, signaling, and pathogen infection), as well as developing novel medicines. The chemoenzymatic synthesis of sialoconjugates *in vitro* can help for the study of the immobilization of multi-structural carbohydrates on a chip surface. The presence of sialic acids in the form of microarray is a high-throughput technology for the rapid analysis of the binding properties of a variety of binding partners including antibodies, pathogens and cells.[164] Using several chemical immobilization methods, diverse sialic acids or their derivatives have been displayed as conjugates on the Gal moiety of glycan acceptors for the screening of SiaT epitope specificities,[165] identification of high affinity Siglec ligands,[166] revealing novel interactions of modified sialic acids with proteins and viruses,[167, 168] and analysis of biointerfaces for protein binding.

In biological systems, multivalency is a valuable strategy to enhance the affinity and specificity of ligand-receptor interactions.[169] The synthesis of multiple ligand clusters, for example in form of dendrimers, can simulate the multivalency affinity, which is helpful for the study of biospecific recognition, inhibition and targeting.[170] Recently, in our group, a “clickable” hybrid nanocluster with cubic symmetry has been synthesized.[169] Owing to its multi-tentacle structure in three-dimensions, it could be an excellent scaffold for sialoside-based dendrimer synthesis. To achieve this aim, the enzymatic synthesis of sialoconjugates needs to be optimized firstly.

## 5.5. Specific aim

- The activity determination of SiaT is usually based on analysis of radioactive[145] or fluorescent labeled probes[146], or by HPLC separation for determining the quantity of formed product,[147] which are all tedious and non-continuous assays. Therefore, a simple sensitive and simple continuous method was highly desirable to be developed for the measurement of kinetic constants and mutagenesis library screening.

- CSS has low activity towards structurally sererely modified sialic acids. Moreover, 2,6SiaT cannot well accept α-galactosides. The newly developed screening method should be instrumental for protein engineering by random mutagenesis in order to extend their substrate range.

- Besides natural sialic acids, the non-natural analogs with modifications at C5, C7, C8 and C9 could be linked to a fluorescent lactose with an α-2,6-linkage.[171] The same typical sialic acids or other derivatives should also be transferred to the same fluorescent Lac with an α-2,3-linkage by 2,3SiaT, or to a fluorescent LacNAc acceptor with α-2,6-linkage by 2,6SiaT.

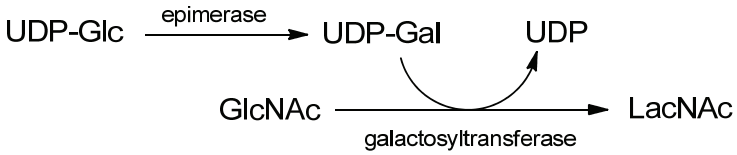
- Many potential SiaT sequences can be found from the genome sequence information in the Genbank database using homology search. With this method, novel bacterial SiaTs should be cloned and identified by genetic engineering.

# 6. Chemoenzymatic Synthesis of *N*-Acetyl-D-lactosamine Precursors

## 6.1. Introduction

*N*-Acetyl-D-lactosamine (LacNAc) exists broadly at the termini of oligosaccharides, such as present in *N*- and *O*-linked glycoproteins and glycolipids that are found in Nature on the cell surface. Additionally, the galactosyl (Gal) and *N*-acetyl-glucosamine (GlcNAc) moieties can be modified forming diverse carbohydrate post-glycosylation modifications (PGMs).[172] In many organisms, a LacNAc moiety performs the role as an acceptor for sialylation.[123] Sialyltransferases transfer the CMP activated sialic acids (CMP-sialic acid) to a LacNAc or lactose (Lac) terminus of oligosaccharides producing α-2,3- or 2,6-sialoconjugates.[123] Therefore, the synthesis of LacNAc precursors is of high importance to study their enzymatic properties of SiaTs, especially the substrate specificity.

The enzymatic synthesis of LacNAc *in vitro* is catalyzed by β-1,4-galactosyltransferase (E.C. 2.4.1.22, GalT), which transfers an UDP activated galactose (UDP-Gal) to the GlcNAc moiety, forming a β-1,4-glycosidic bond, (Scheme 6.1) in analogy to the natural synthesis pathway of LacNAc *in vivo*.[173] GalT is present broadly in mammalian organisms. However, the activity of mammalian GalT requires α-lactalbumin for Glc based acceptors,[174] which is inconvenient for the synthesis of LacNAc and Lac precursors. Also, mammalian GalTs cannot be well expressed in soluble form in recombinant bacteria even after removal of their membrane insertion segments.[175] On the other hand, soluble bacterial GalT can be expressed well in *E. coli* and can perform the reaction without α-lactalbumin activation, which therefore becomes the ideal biocatalyst for the galactosylation *in vitro*. Up to now, several useful bacterial GalTs have been found in *Neisseria meningitides* ($GalT_{nme}$)[175, 176], *Neisseria gonorrhoeae* ($GalT_{ngo}$)[176] and *Helicobacter pylori* ($GalT_{hpy}$)[176, 177], and the GalTs from these three organisms have been successfully cloned and expressed in *E. coli*.[175-178]




Scheme 6.1. Enzymatic synthesis pathway for *N*-Acetyl-D-lactosamine and its derivatives

GalT from *Neisseria* sp. accommodates a broad substrate range. It can accept not only natural Glc, GlcNAc and Fuc-1,3-GlcNAc as Gal acceptors to produce large-scale LacNAc and Lewis$^x$ tetrasaccharides, but also unnatural Glc and GlcNAc moieties carrying modification at C1, C2 and C6.[179-184] GalT from *Helicobacter pylori* ($GalT_{hpy}$) has been shown to have only $3\times10^{-4}$ lower activity towards Glc than that towards GlcNAc.[177] It also is unable to accept any Glc derivatives and even GlcNAc with α-anomeric conjugation.[184] However, Namdjou *et al.* also reported that $GalT_{hpy}$ has some activity towards 4-nitrophenyl labeled Glc (*p*NP-Glc),[185] and can also accept 4-nitrophenyl labeled mannose (*p*NP-Man) with very low activity.[185] Therefore, $GalT_{hpy}$ has been used for thioglycoside and poly-LacNAc synthesis.[185, 186]

UDP-Gal is essential for galactosyltransferation as the only acceptable donor for GalT. However, UDP-Gal is too expensive to be used for large scale synthesis (€473.5/100mg, Sigma-Aldrich, 2012). To deal with this problem, a practicable solution is using UDP-galactose 4-epimerase (E.C. 5.1.3.2, GalE) to transform UDP-glucose (UDP-Glc) into UDP-Gal, because UDP-Glc (€71.8/100mg, Sigma-Aldrich, 2012) is much cheaper than UDP-Gal. Thus, this could be a relative economical pathway for LacNAc synthesis (Figure 6.1). The GalE enzyme from *E. coli* ($GalE_{eco}$) has been cloned.[187] It can be added to the synthesis reaction as coupled enzyme or even in combination with UTP-glucose-1-phosphate uridylyltransferase (EC 2.7.7.9, GalU) for UDP-Glc preparation,[184] but also was expressed as fusion protein[188] or coexpressed with GalT for whole cell catalysis.[180-182, 189]

As natural acceptor for sialyltransferases, LacNAc precursors have significant research value for the synthsis of neo-sialoconjugates *in vitro*. In this chapter, $GalT_{hpy}$ and $GalE_{eco}$ have been cloned and expressed in *E. coli*, followed by the chemoenzymatic preparation of LacNAc and the fluorophore-labeled LacNAc and Lac precursors for a further synthesis of neo-sialoconjugates.

## 6.2. Results and discussion

### 6.2.1. Cloning, expression and characterization of β-1,4-galactosyltransferase from *Helicobacter pylori*

The gene of β-1,4-galactosyltransferase from *Helicobacter pylori* ($GalT_{hpy}$) and cloning work have been reported.[177, 190] The amino acid sequence of $GalT_{hpy}$ consists of 274 residuces with molecular mass of 31 kDa, but the enzyme has no significant similarity to the GalT from *Neisseria* sp. which have been researched widely.[177] Due to the pathogenicity of *Helicobacter pylori*, it was decided to use an artificial gene of $GalT_{hpy}$ for its cloning in this work. The gene of $GalT_{hpy}$ was synthesized commercially and cloned into Tuner BL21 host cells using the pET19b vector for protein over-expression (Figure 6.1a). The recombinant protein expression was induced by 0.1 mM IPTG at 28 °C overnight. The SDS-PAGE analysis displaying the expression level is shown in Figure 6.1b. It seems that in the crude cell extract all of recombinant $GalT_{hpy}$ existed in the pellet. To improve the soluble expression level, the culture conditions including growth temperature (25°C-37°C), $OD_{600}$ for induction value (0.3-1.0) and concentration of IPTG (0.01-1 mM) were optimized, but no significant improvement could be observed. However, fortunately, quite strong GalT activity could be found both in the crude extract and supernatant of the recombinant cells (Figure 6.1c), even if $GalT_{hpy}$ could hardly be detected in SDS-PAGE gels.

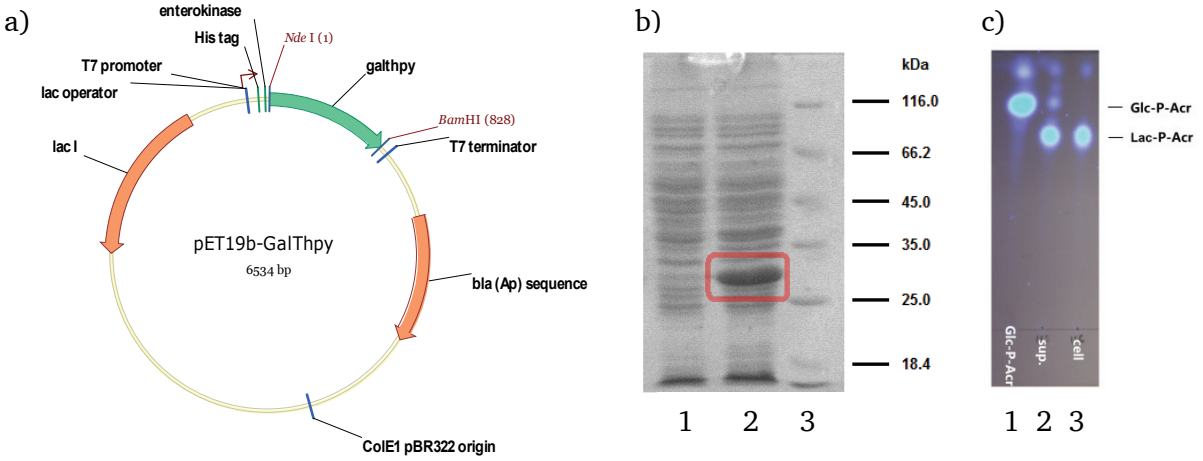



Figure 6.1. Cloning, expression and activity determination of GalT from *Helicobacter pylori*. a) Plasmid map of the expression plamid pET19b-$GalT_{hpy}$. b) Expression level of the recombinant $GalT_{hpy}$ displayed by SDS-PAGE. 1: Supernatant of crude cell extract. 2: Crude cell extract. 3: Marker c) Activity determination of $GalT_{hpy}$ by fluorescence labeled acceptor (Glc-D-T-P-Acr, see structure in scheme 6.2). The reaction was carried out in 100 µl scale with 20 mM Tris-HCl (pH 7.5) and 5 mM $MnCl_2$. The reaction system contains 3 mM Glc-D-T-P-Acr, 5 mM UDP-Gal and 50 µl enzyme sample.1: Fluorescence labeled acceptor Glc-D-T-P-Acr. 2: Supernatant of crude cell extract. 3: Crude cell extract.

The low soluble expression of the protein of interest in the *E. coli* system is mainly caused by two reasons. One is that the over-expressed protein cannot be folded in a correct way during its expression, and consequently most of the expressed protein agglomerates in the form of inclusion bodies. The other reason may be due to a hydrophobic trans-membrane anchoring of the recombinant protein, which causes this kind of protein to coprecipitate with the membrane fragments during the cell lysis. Similar difficulty in the $GalT_{hpy}$'s soluble expression was also reported in the literature.[177, 184, 190] Lau *et al.* tried to express $GalT_{hpy}$ as a fusion protein with an N-terminal maltose binding protein (MBP)-tag and C-$His_6$-tag.[184] In this work, two alternative ways, chaperone co-expression and gene modification, were used for its expression optimization.

*In vivo*, the protein folding has been shown to be an energy-dependent process mediated by molecular chaperones and foldases.[191] The former binds partially folded proteins and maintains them in a soluble or translocation-competent conformation, while the latter can accelerate the rate-limiting steps along the folding pathway, for example the isomerization of peptidyl-prolyl bonds and the formation and reshuffling of disulfide bridges.[191] As useful tools for protein expression, the chaperonins DnaK-DnaJ-GrpE, GroEL-GroES and Tf in *E. coli* have been well identified and successfully applied for heterologous protein expression, promoting the soluble expression level observably.[192, 193] Therefore, in order to improve the soluble expression level of $GalT_{hpy}$ in *E. coli* cells, it was coexpressed with the chaperonins GroEL-GroES and DnaK-DnaJ-GrpE, respectively. All of these chaperones have a high expression level in *E. coli*. Unfortunately, no significantly increased soluble expression of recombinant $GalT_{hpy}$ was found in SDS-PAGE gel (Figure 6.2).

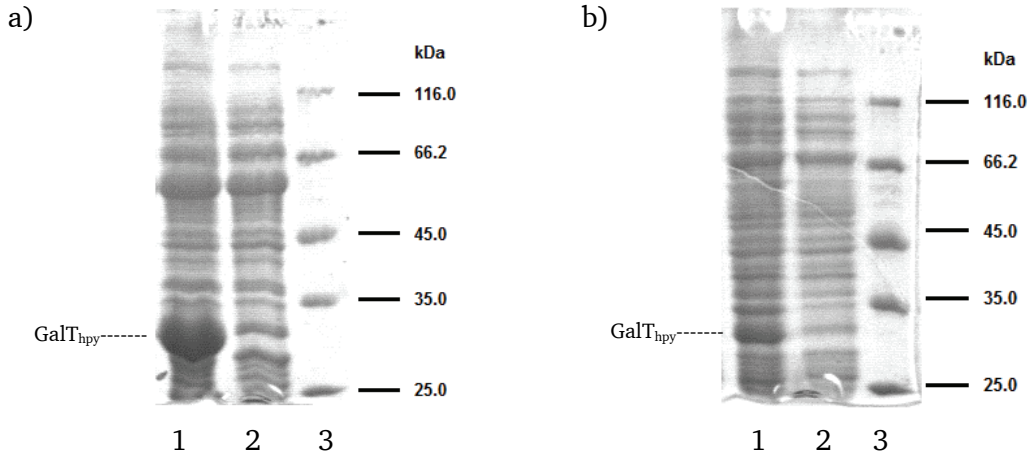



Figure 6.2. Coexpression of GalT from *Helicobacter pylori* with chaperonins a) GroEL-GroES b) DnaK-DnaJ-GrpE. 1: Crude cell extract. 2: Supernatant of crude cell extract. 3: Marker.

The failing of the chaperone co-expression but nevertheless high GalT activity identified both in supernatant and pellet of the recombinant cell lysate indicated that $GalT_{hpy}$ might be a membrane protein. To analyze whether $GalT_{hpy}$ contains a transmembrane fragment, the protein sequence of $GalT_{hpy}$ was aligned to that of fucosyltransferase from *Helicobacter pylori* ($FucT_{hpy}$), for which it has been verified that it carries a transmembrane fragment at the C-terminus.[194] The alignment result for these two protein sequences is shown in Sequence 6.1. The transmembrane fragment of proteins has a typical characteristic: Lys and Arg are repeated with a fixed interval that can fold into a putative amphipathic α-helix structure in the protein's tertiary structure which functions as a membrane anchor. In the FucT case, the deletion of its transmembrane fragment increased both expression level and solubility of $FucT_{hpy}$, without significant loss of the activity.[194, 195] At the C-terminus, $GalT_{hpy}$ also contains a relatively high percentage of Lys and Arg. Thus, this 27 residues transmembrane-like fragment of $GalT_{hpy}$ was deleted to improve the solubility of $GalT_{hpy}$ in *E. coli*. To avoid a negative influence of the His-tag, two different pET vectors (pET19b and pET21a) were used for the cloning, which add His-tag on either the C- or N-terminus of $GalT_{hpy}$, respectively. However, neither of the expressed modified $GalT_{hpy}$ variants became sufficiently soluble. As a result, the crude extract of recombinant cells containing unpurified $GalT_{hpy}$ was used for the subsequent synthesis work.

```
                1                                                                    75
C-FucThpy     (1) ------------------------------------------------------------------------
  GalThpy     (1) MLRVFIISLNQKVCDTFGLVFRDTTTLLNNINATHHQAQIFDAIYSKTFEGGLHPLVKKHLHPYFITQNIKDMGI
               76
                         150
C-FucThpy     (1) ------------------------------------------------------------------------
  GalThpy    (76) TTNLISEVSKFYYALKYHAKFMSLGELGCYASHYSLWEKCIELNEAICILEDDITLKEDFKEGLDFLEKHIQELG
              151
                       225
C-FucThpy     (1) ------------------------------------------------------------------------
  GalThpy   (151) YVRLMHLLYDPNVKSEPLNHKNHEIQERVGIIKAYSHGVGTQGYVITPKIAKVFKKHSRKWVVPVDTIMDATFIH
              226                                                              275
C-FucThpy     (1) --------------------FKIYRKAYQKSLPLLRTIRRWVKK------
  GalThpy   (226) GVKNLVLQPFVIADDEQISTIARKEEPYSPKTALMREIHFKYLKYWQFV-
```

Sequence 6.1. Alignment of the protein sequence of $GalT_{hpy}$ with the C-terminus of $FucT_{hpy}$. Lys and Arg residues are marked with acid blue. The transmembrane-like fragment, which contains 27 amino acid residues, is marked with red color.

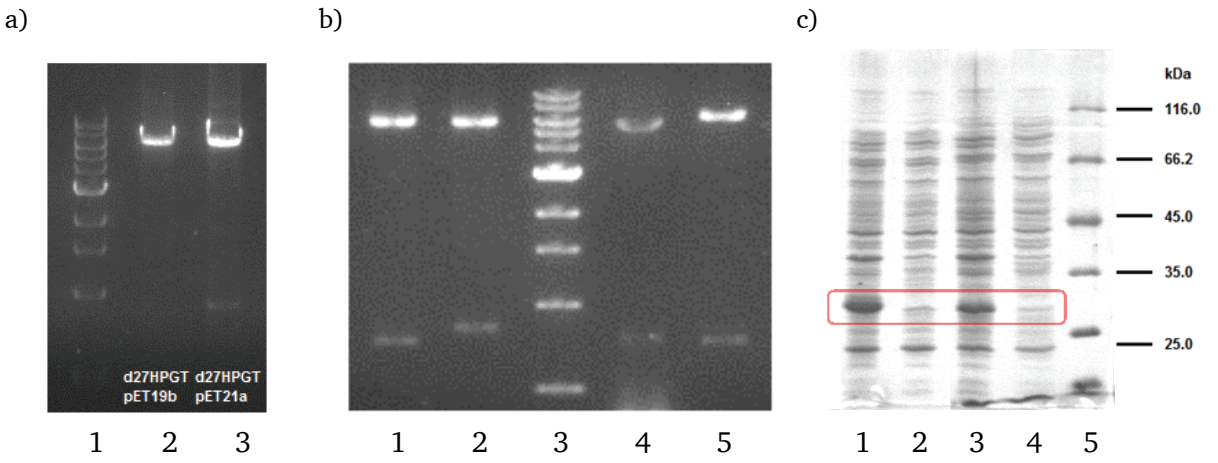



Figure 6.3. Deletion of the transmembrane-like fragment at the C-terminus of $GalT_{hpy}$. a) Identificaiton of PCR product for the deletion. 1: Marker. 2: PCR product of pET19b-d27$GalT_{hpy}$. 3: PCR product of pET21a-d27$GalT_{hpy}$. b) Restriction identification of the plasmids (*Nde* I and *BamH* I). 1: pET21a-d27$GalT_{hpy}$. 2: pET21a-$GalT_{hpy}$. 3: Marker. 4: pET19b-$GalT_{hpy}$. 5: pET19b-d27$GalT_{hpy}$. c) Idendification of expression level by SDS-PAGE. 1: d27$GalT_{hpy}$-C-His, crude cell extract. 2: d27$GalT_{hpy}$-C-His, supernatant of crude cell extract. 3: d27$GalT_{hpy}$-N-His, crude cell extract. 4: d27$GalT_{hpy}$-N-His, supernatant of crude cell extract. 5: Marker.

### 6.2.2. Cloning, expression and characterization of UDP-galactose 4-epimerase from *Escherichia coli*

UDP-galactose 4-epimerase (E.C. 5.1.3.2, GalE) catalyzes the interconversion of UDP-Glc and UDP-Gal *in vivo*. The GalE from *E. coli* ($GalE_{eco}$) is a homodimer (79kDa), in which each subunit contains 338 amino acid residues [187]. The gene of $GalE_{eco}$ was obtained from the chromosomal DNA of *E. coli* strain K-12 using PCR, cloned into pET19b vector and expressed in Tuner BL21 cells. The expressed recombinant $GalE_{eco}$ contains a $His_6$-tag at the C-terminus. Because this GalE was homologously cloned from *E. coli*, a high soluble expression level could be achieved (Figure 6.4). The recombinant enzyme was purified by Ni-affinity chromatography.

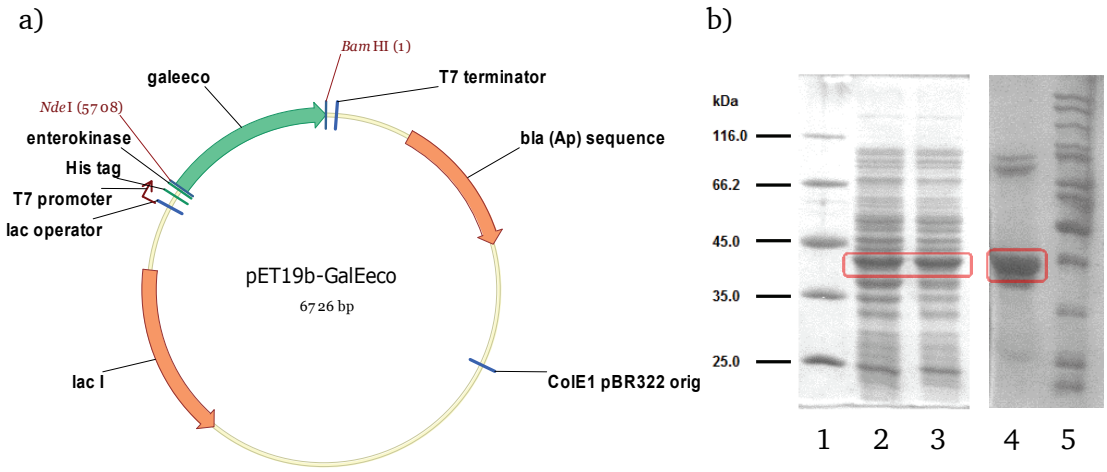
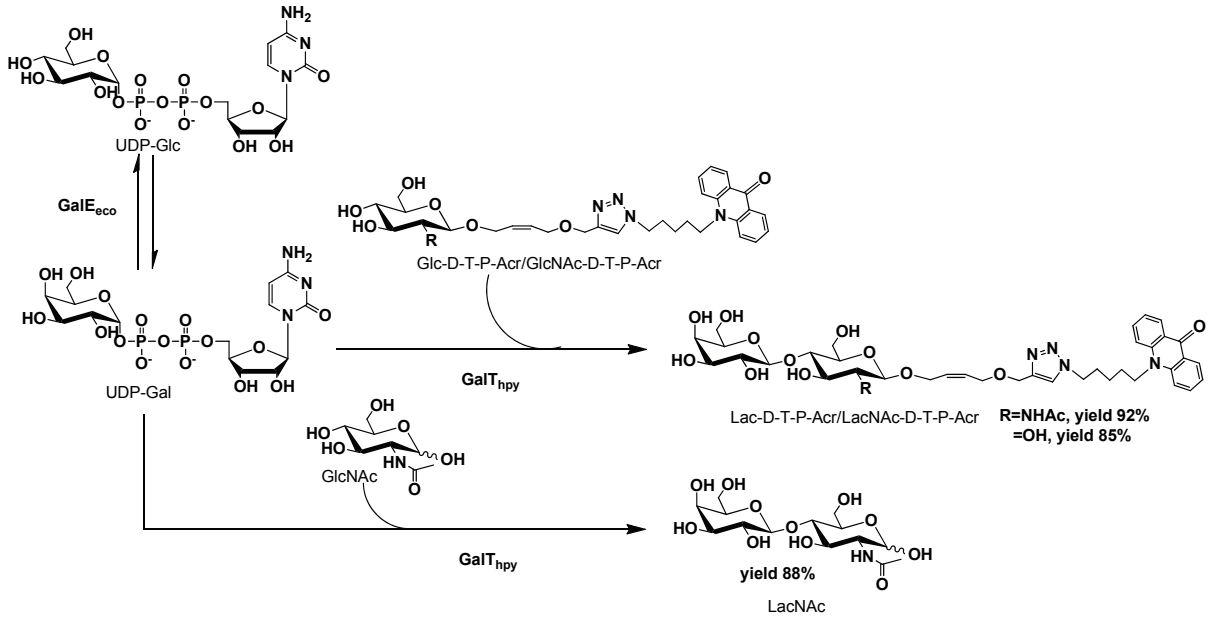



Figure 6.4. Cloning and expression of GalE from *E. coli*. a) Plasmid map of the expression plamid pET19b-GalE$_{eco}$. b) Expression level of the recombinant GalE$_{eco}$ displayed by SDS-PAGE. 1: Marker. 2: Crude cell extract. 3: Supernatant of crude cell extract. 4: Purified GalE$_{eco}$ by Ni-affinity chromatography. 5: Marker.

### 6.2.3. Chemoenzymatic synthesis of *N*-acetyl-D-lactosamine precursors

Using recombinant GalE$_{eco}$ and GalT$_{hpy}$, the synthesis of LacNAc and its fluorophore-labeled derivative was carried out as one-pot reactions, in which UDP-Glc was converted into UDP-Gal which is subsequently consumed as glycosyl donor, while GlcNAc or fluorophor-labeled GlcNAc played as acceptors. Although GalE$_{eco}$ catalyzes a fully reversible reaction with only 25% final conversion of UDP-Glc to UDP-Gal,[187] the galactosylation consumed UDP-Gal and promoted the continuous further production of UDP-Gal. GalT$_{hpy}$ has a relative high activity with GlcNAc and its derivatives. The total conversion of LacNAc and LacNAc-D-T-P-Acr thus reached virtually 100% as judged from the TLC monitoring of the reaction. The final isolated yields of these two compounds at the 100 mg scale were as high as 88% and 92%, respectively (Scheme 6.2). The product LacNAc-D-T-P-Acr was subsequently used as acceptors for the synthesis of neo-sialoconjugates in Chapter 10.




Scheme 6.2. Chemoenzymatic synthesis of *N*-Acetyl-D-lactosamine and lactosamine precursors

GlcNAc-α-ethynyl, GlcNAc-β-ethynyl, Glc-β-ethynyl glycosides and free Glc were also tested as acceptors to confirm the results reported by Lau *et al.*, who found that $GalT_{hpy}$ could not accept any Glc and α-GlcNAc derivatives.[173] However, in our TLC determination, GlcNAc-α-ethynyl glycoside was also well accepted by $GalT_{hpy}$. Interestingly, the recombinant $GalT_{hpy}$ could also use Glc-β-ethynyl glycoside and Glc as an acceptor, though its activity was lower than that towards GlcNAc and its derivatives (Figure 6.5). Lau *et al.*'s recombinant $GalT_{hpy}$ was expressed in the fusion form with MBP and $His_6$ tags at N and C terminus, respectively, in order to improve its soluble expression.[173] Conversely, the recombinant $GalT_{hpy}$ described here contained only a $His_6$-tag at the N terminus. There is the possibility that the fusion tag at the C-terminus or at both N- and C-termini affects the specificity of $GalT_{hpy}$. It is a pity that the crystal structure of $GalT_{hpy}$ has not been determined yet, and also that there is no similar enzyme with high homology toward $GalT_{hpy}$ that could be used as template for a 3D structure simulation. Therefore, up to now, it is still difficult to evaluate an effect of the MBP tag on the tertiary structure of $GalT_{hpy}$. To confirm the Lac derivative product catalyzed by $GalT_{hpy}$, Lac-D-T-P-Acr was synthesized from Glc-D-T-P-Acr (50 mg scale, yield 85%), which reaction performed with significantly lower rate than that using GlcNAc-D-T-P-Acr as acceptor. Although the chemical preparation of Lac-D-T-P-Acr from Lac as starting material is much more economic,[196] the enzymatic galacosylation offers an alternative route for its synthesis.

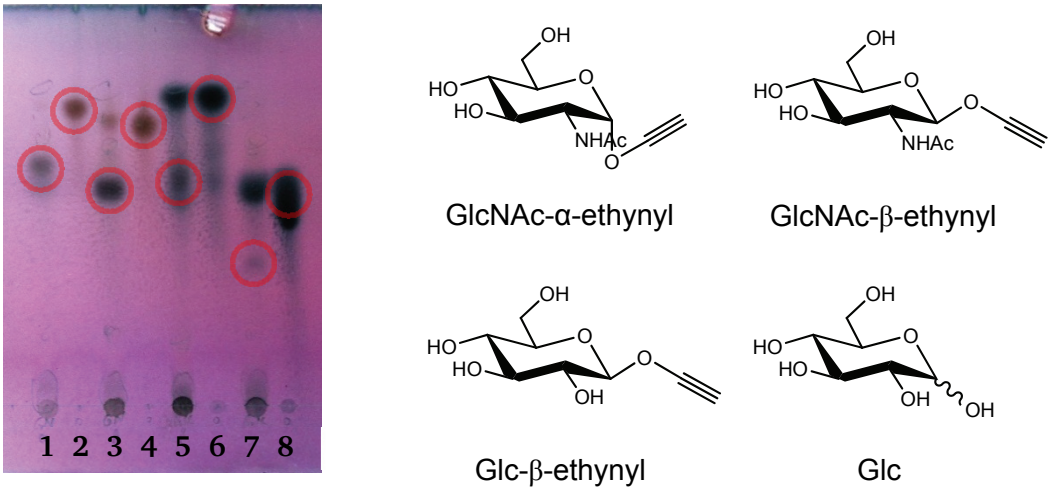



Figure 6.5. TLC identification of the specificity of $GalT_{hpy}$. 1: Product from GlcNAc-α-ethynyl glycoside. 2: GlcNAc-α-ethynyl glycoside. 3: Product from GlcNAc-β-ethynyl glycoside. 4: GlcNAc-β-ethynyl. 5: Product from Glc-β-ethynyl glycoside. 6: Glc-β-ethynyl glycosides. 7: Product from Glc. 8: Glc.

## 6.3. Conclusion

The LacNAc moiety at the terminus of oligosaccharides is an important natural substrate of SiaT. Synthesis of LacNAc and its derivatives *in vitro* can extend the diversity of acceptors for the synthesis of neo-sialoconjugates. In this chapter, the β-1,4-galactosyltransferase from *Helicobacter pylori* has been cloned and over-expressed in an *E. coli* expression system using a synthetic gene sequence. The expression of $GalT_{hpy}$ has been optimized with the aid of chaperone co-expression and gene truncation. Although the expressed recombinant $GalT_{hpy}$ remained largely insoluble, high GalT activity could nevertheless be identified in the crude extract of recombinant cells. The GalE from *E. coli* was also cloned for the interconversion of UDP-Glc to UDP-Gal to reduce the substrate cost of the galactosyltransfer reaction. Finally, free LacNAc and LacNAc-D-T-P-Acr, as well as Lac-D-T-P-Acr have been synthesized from UDP-Glc donor and using GalNAc, GalNAc-D-T-P-Acr and Gal-D-T-P-Acr as acceptor substrates, with isolated yields of 88%, 92% and 85%, respectively, on a 100 mg scale each for further preparative utilization.

# 7. Cloning, Expression and Characterization of α-2,3- and α-2,6-Sialyltransferases from *Photobacterium* sp.

## 7.1. Introduction

Bacterial α-2,3/2,6-sialyltransferases (2,3/6SiaTs) are classified into the GT80 family (www.cazy.org) which currently contains 15 different SiaT sequences from *Alistipes shahii* (GenBank accession number: CBK63974.1), *Haemophilus ducreyi* (GenBank accession number: AAP95068.1), *Pasteurella multocida* (GenBank accession number: AET16017.1,[197] AAY89061.1,[147] AFI46264.1, AFF24354.1[198] and AAK02272.1[199]), *Photobacterium* sp. (GenBank accession number: BAA25316.1,[145] BAI49484.1,[200] BAF91416.1,[160] BAF63530.1,[161] BAF92025.1[162] and BAF92026.1[162]), *Shewanella piezotolerans* (GenBank accession number: ACJ31674.1[201]), and *Vibrio* sp. (GenBank accession number: BAF91160.1[163]), respectively. Among theese, the enzymatic properties of a multifunctional SiaT from *Pasteurella multocida* (2,3/6SiaT$_{pmu}$) has been studied intensively, and the enzyme has been applied for the synthesis of sialoside libraries.[137, 147, 150, 202-204] Recently, the discovery of 2,3- and 2,6SiaTs in marine bacteria from *Photobacterium* sp. with high sialyltransfer activity has immensely expanded the membership size of the bacterial SiaT family.[145, 158, 160-163, 200] Some of these enzymes have unique properties. For example, the 2,6SiaT from *Photobacterium leiognathi* JT-SHIZ-145 (2,6SiaT$_{ple}$) has an optimum at pH 8 instead of 6 to 6.5 reported for other SiaTs;[160] an alkaline optimum is more favorable for sialoconjugate synthesis, because CMP-activated sialic acids are un stable under acidic conditions. The synthetic potential of the SiaTs from *Photobacterium* sp. has also been proved. Especially, a 2,6SiaT from *Photobacterium damselae* JT0160 (2,6SiaT$_{pda}$) has been successfully applied to the synthesis of natural and non-natural sialosides.[144, 202, 205] As a result, 2,3SiaT and 2,6SiaT from *Photobacterium* sp. became the preferred candidates in this thesis for the purpose of synthesis of neo-2,3- and 2,6-sialocojugates.

For synthetic applications, an accurate study of the substrate specificity and kinetics of SiaT is quite significant. Because of the difficulty of the oligosaccharide detection and quantification, the assay methods used for SiaT activity monitoring employ radioactive [145] or fluorescently labeled CMP-Neu5Ac[146], or determine the formation of product directly by HPLC.[147]

However, none of these methods is easy to be performed. The radio-labeled compound is expensive and harmful to health, and requires a special detector. The modification of either donor or acceptor may change the activity of SiaT or limit the variety of accessible substrates, which will narrow the scope for the substrate specificity determination of SiaT. Therefore, a sensitive, simple and rapid screening assay method is still required for the activity determination towards a broader substrate variation.

In this chapter, the 2,3-SiaT from *Photobacterium phosphoreum* JT-ISH-467 ($2,3SiaT_{pph}$) and 2,6-SiaT from *Photobacterium leiognathi* JT-SHIZ-145 ($2,6SiaT_{ple}$) were cloned, over-expressed in *E. coli* and purified by affinity chromatography. Their substrate tolerance was determined with a novel pH-based assay method.

## 7.2. Results and discussion

### 7.2.1. Cloning, expression and characterization of α/β-galactoside α-2,3-sialyltransferase from *Photobacterium phosphoreum*

The α/β-galactoside α-2,3-sialyltransferase from *Photobacterium phosphoreum* ($2,3SiaT_{pph}$) is a novel 2,3SiaT that was reported in 2007,[161] for which recently also a protein crystal structure has been determined [148]. Therefore, it can be used as a standard enzyme and structure template for the research of other 2,3SiaTs in the same family. The sequence of $2,3SiaT_{pph}$ has 409 amino acid residues, among which the first 24 residues at its N-terminus have been identified as a signal sequence.[159] The gene of Δ24-$2,3SiaT_{pph}$ was synthesized commercially and ligated into the pET19b vector (Figure 7.1), then transformed into BL21(DE3) host cells for over-expression. The SDS-PAGE analysis shows that truncated the recombinant $2,3SiaT_{pph}$ is well expressed in the soluble form by the *E. coli* expression system (Figure 7.2a). However, activity identification using the fluorolabeled lactose (Lac-D-T-P-Acr) as acceptor revealed that besides high 2,3SiaT activity from the protein of interest, the crude extract sample also contained β-galactosidase activity (Figure 7.1b). Since $2,3SiaT_{pph}$ should not have any hydrolytic activity, the observed galactosidase activity probably came from the host strain BL21(DE3). Therefore, the expression host was changed into BL21Tuner, in which galactosidase activity was indeed practically absent upon host cell screening. The recombinant enzyme was purified using Ni-affinity chromatography (Figure 7.1c).

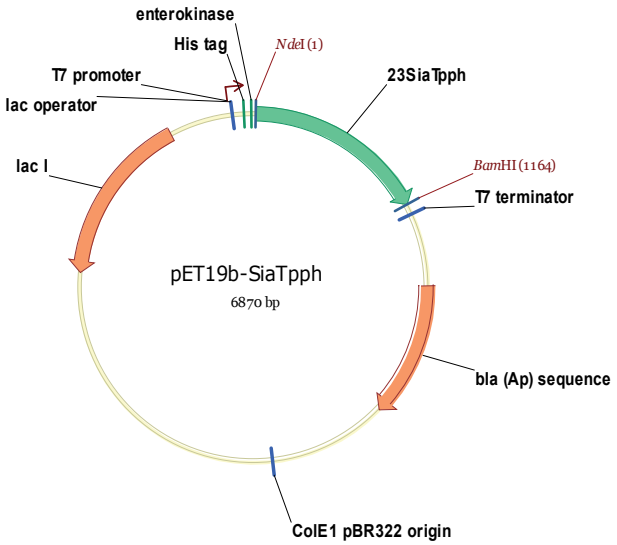



Figure 7.1. Plasmid map of the expression plamid pET19b-SiaT$_{pph}$.

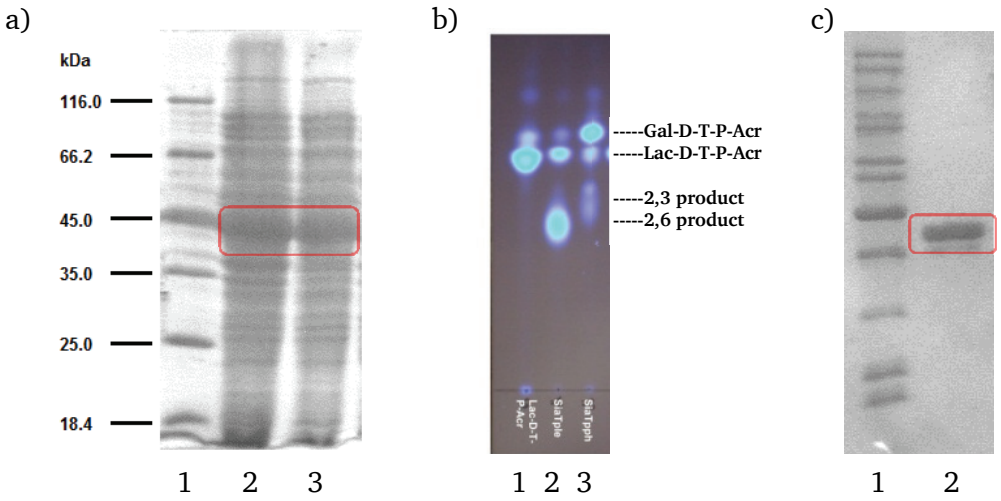



Figure 7.2. Cloning, expression and activity determination of 2,3SiaT from *Photobacterium phosphoreum*. a) Expression level of the recombinant SiaT$_{pph}$ in BL21(DE3) displayed by SDS-PAGE. 1: Marker. 2: Crude cell extract. 3: Supernatant of crude cell extract. b) Activity determination of 2,3SiaT$_{pph}$ by fluorescent labeled acceptor (Lac-D-T-P-Acr). The reaction was carried out in 100 µl scale with 20 mM Tris-HCl (pH 8). The reaction system contains 3 mM Lac-D-T-P-Acr (see scheme 6.2), 5 mM CMP-Neu5Ac and 50 µl enzyme sample.1: Fluorescence labeled acceptor Lac-D-T-P-Acr. 2: 2,6-Sialylation reference. 3: Rection with supernatant of crude cell extract. c) Purified 2,3SiaT$_{pph}$ by Ni-affinity chromatography. 1: Marker. 2: Purified 2,3SiaT$_{pph}$.

### 7.2.2. Cloning, expression and characterization of β-galactoside α-2,6-sialyltransferase from *Photobacterium leiognathi* JT-SHIZ-145

The β-galactoside α-2,6-sialyltransferase from *Photobacterium leiognathi* JT-SHIZ-145 (2,6SiaT$_{ple}$), which was isolated from the gut of a Japanese barracuda, is produced from an

open reading frame of 1,494 base pairs encoding a predicted protein of 497 amino acid residues.[160] The deduced amino acid sequence of 2,6SiaT$_{ple}$ showed 68.2% homology to 2,6SiaT$_{pdm}$. In the previous study, the optimum pH of SiaT$_{ple}$ was determined at 8. Its transfer ability for mono- and di-saccharides and glycoproteins as sialyl acceptors, and the kinetic parameters of a truncated version of the sialyltransferase have been studied previously.[160]

In the present study, the DNA sequence of the 2,6SiaT$_{ple}$ gene was optimized for expression in *E. coli* and synthesized commercially. For the 2,6SiaT$_{pdm}$, it was reported that the first 15 amino acid residues at the N-terminus is a signal sequence and unrelated to its activity. [159] A similar sequence was also found in the N-terminus of 2,6SiaT$_{ple}$[160] and thus, the first 15 amino acid residues of 2,6SiaT$_{ple}$ were similarly deleted. The truncated 2,6SiaT$_{ple}$ gene fragment was ligated into the pET19b vector to construct the expression plasmid, which was then transformed into BL21 Tuner competent cells. The map and identification of the expression plasmid (pET19b-SiaT$_{ple}$) are shown in Figure 7.3. The 2,6SiaT$_{ple}$ could also be expressed well in soluble form in the *E. coli* system (Figure 7.4a). The purity of the recombinant 2,6SiaT$_{ple}$ purified by Ni-affinity chromatography is displayed in Figure 7.4b. The recombinant 2,6SiaT$_{ple}$ contains a His$_6$-tag at the N-terminus, which gives higher activity than when produced with a C-terminal His-tag.[159] For a rapid activity test of the recombinant 2,6SiaT$_{ple}$ also the fluorescence labeled lactose (Lac-D-T-P-Acr) was used as the acceptor together with the lysis supernatant of the recombinant strain as biocatalyst. The TCL result indicated a high activity of recombinant 2,6SiaT$_{ple}$ (Figure 7.4c).

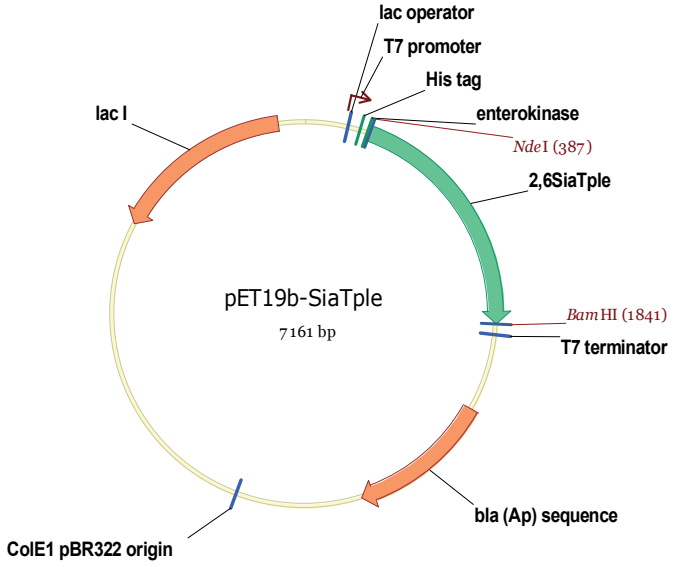



Figure 7.3. Plasmid map of the expression plamid pET19b-SiaT$_{ple}$.

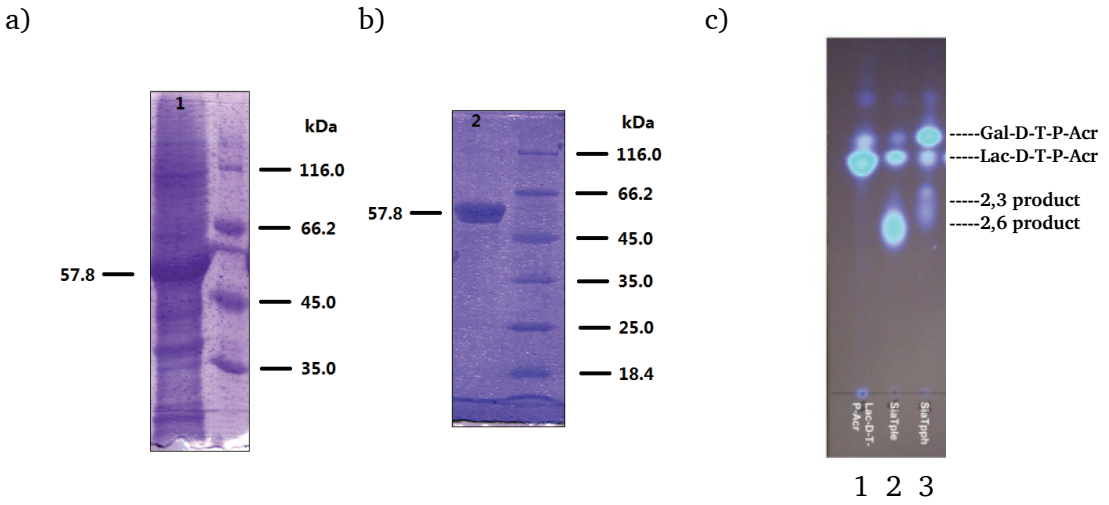



Figure. 7.4. Expression, purification and activity identification of the recombinant 2,6SiaT$_{ple}$. a) Expression level of 2,6SiaT$_{ple}$ in Tuner BL21. b) Purified recombinant 2,6SiaT$_{ple}$. c) Activity determination of 2,6SiaT$_{ple}$ by fluorescent labeled acceptor (Lac-D-T-P-Acr). The reaction was carried out in 100 µl scale with 20 mM Tris-HCl (pH 8). The reaction system contains 3 mM Lac-D-T-P-Acr (see scheme 6.2), 5 mM CMP-Neu5Ac and 50 µl enzyme sample. 1: Fluorescence labeled acceptor Lac-D-T-P-Acr. 2: Rection with supernatant of crude cell extract. 3: 2,3-Sialylation reference.

### 7.2.3. Identification of regioselectivities of 2,3SiaT$_{pph}$ and 2,6SiaT$_{ple}$ by synthesis of Neu5Ac-2,3-Lac-D-T-P-Acr and Neu5Ac-2,6-Lac-D-T-P-Acr

The high sialyltransfer activities of 2,3SiaT$_{pph}$ and 2,6SiaT$_{ple}$ have been proved from the rapid activity test using Lac-D-T-P-Acr as acceptor. In order to confirm the α-2,3- or α-2,6-linkage in each product, the sialylation of Lac-D-T-P-Acr was performed in a 20 mg scale. Both of the 2,3- and 2,6-sialylated products were isolated and analyzed by NMR spectrum. The structure variance between the products Neu5Ac-2,3-Lac-D-T-P-Acr and Neu5Ac-2,6-Lac-D-T-P-Acr is the different sialylation positions. In Neu5Ac-2,3-Lac-D-T-P-Acr, the sialylation at OH3" position is accompanied by a large deshielding for the corresponding glyconic hydrogen.[206] Therefore, the corresponding peaks if H3" in $^{1}$H spectrum shift from around 3.6 ppm in Neu5Ac-2,6-Lac-D-T-P-Acr to 4.1 ppm in Neu5Ac-2,3-Lac-D-T-P-Acr (Figure 7.5). This large deshielding (around 0.5-0.6 ppm) can be a main reference to distinguish 2,3-sialosides and 2,6-sialosides.

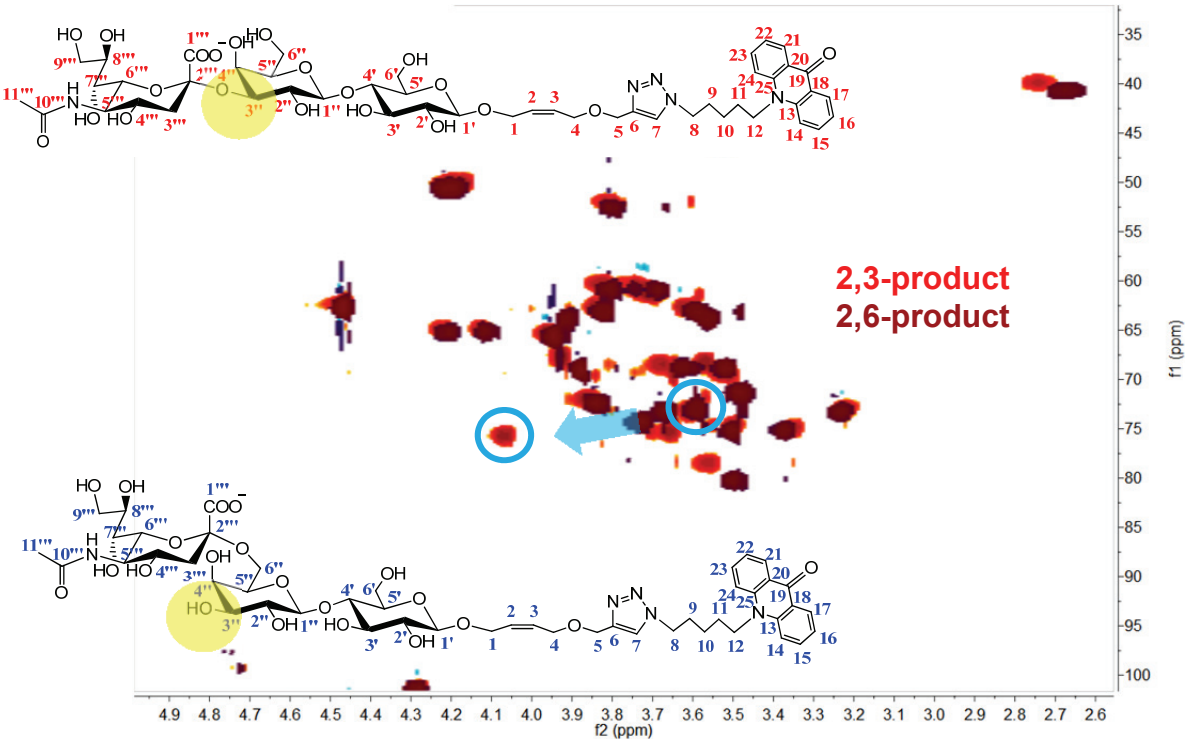
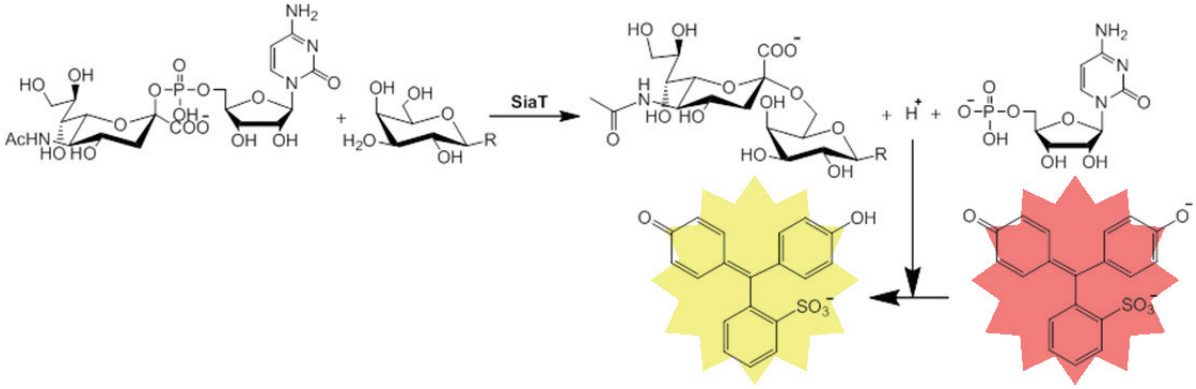



Figure 7.5. NMR spectrum (gs-HSQC) comparison of Neu5Ac-2,3-Lac-D-T-P-Acr and Neu5Ac-2,6-Lac-D-T-P-Acr. In Neu5Ac-2,3-Lac-D-T-P-Acr, the sialylation at OH3'' position is accompanied by a large deshielding for the corresponding glyconic hydrogen. The corresponding peaks if H3'' in $^{1}$H spectrum shift from around 3.6 ppm in Neu5Ac-2,6-Lac-D-T-P-Acr to 4.1 ppm in Neu5Ac-2,3-Lac-D-T-P-Acr.

### 7.2.4. Development of a pH-based assay for sialyltransferases

For the specificity determination with a number of potential substrates, a fast and sensitive method is required. Recently, the principle pH-based assay has been introduced successfully for the activity determination of glycosyltransferases, such as GalT and GalNAcT.[207, 208] In principle, any enzymatic reaction using substrate/product pairs that have sufficiently different p$K_a$ values and that therefore cause changes in pH upon conversion can be evaluated qualitatively and quantitatively by adding a suitable pH indicator and monitoring OD variations during the reaction in the absence of a buffer system. During the sialyltrasfer, the production of a proton upon CMP release decreases the pH value of the reaction system. This pH change could also be well revealed by adding a pH indicator. According to this principle, a pH-based method was developed, in which phenol red (p$K_a$ 8.0) was used as the pH indicator. Its color exhibits a gradual transition from yellow to red over the pH range 6.8 to 8.2, which includes the pH optimum of the enzymatic reaction system (pH 8.0). As discussed in detail in

Chapter 3, a low concentration (1 mM to 3 mM) is necessary for high sensitivity of the pH based assay and a buffer with a p$K_a$ as close as possible to the p$K_a$ of the indicator gives the results.[209] Tris has a p$K_a$ at 8.1 and shows chemical reactivity with any other component present in the sialylation system, which is perfect for its application in this measurement method. Thus, 2 mM Tris-HCl (pH 8.0) was used to slightly buffer the assay.




Scheme 7.1. Principle of the pH-based assay for sialyltransferase analysis

The proton used for quantification is released during the sialylation reaction. Therefore, a standard curve was established for the linear relationship between proton concentration and absorbance of phenol red at 560 nm using 0 mM to 0.5 mM HCl (Figure 7.6a). Because of the relatively higher acidity of phosphate, the effect of stoichiometric released protons on the pH value is much higher than that of bicarbonate released in the TK assay. Therefore, the LOD and LOQ of this standard curve can reach 0.014 mM and 0.041 mM. Using a few substrates as acceptors for the primary test, the time curves of absorbance (Figure 7.6b) clearly revealed the different activities of SiaT towards the varying acceptors, including α- and β-configurated galactose derivatives. Absorbance data for the first 60 s were recorded and converted into concentration of protons according to the standard curve. A typical time curve of concentration of proton using lactose as acceptor is shown in Figure 7.6c. The slope of this curve was used to calculate the specific activity of SiaT according to the Formula 7.1. One unit was defined as the amount of enzyme required to produce 1 $\mu$mol of protons per minute.

$$specific\ activity\ \mathbf{U/mg} = \frac{slope\ \mathbf{mM/s} \times 10^3 \times 60\mathbf{s} \times 200 \times 10^{-6}\mathbf{L}}{amount\ of\ enzyme\ \mathbf{mg}} \quad \text{(Formula 7.1)}$$

Variation of the amount of SiaT for an equal increment of OD change confirmed the linear relationship between activity and amount of enzyme (Figure 7.6d), which demonstrates that this method can be used for the quantitative determination of the absolute and relative specific activity of different SiaT. The observable LOD and LOQ values are 0.6 mU and 1.9 mU, respectively.

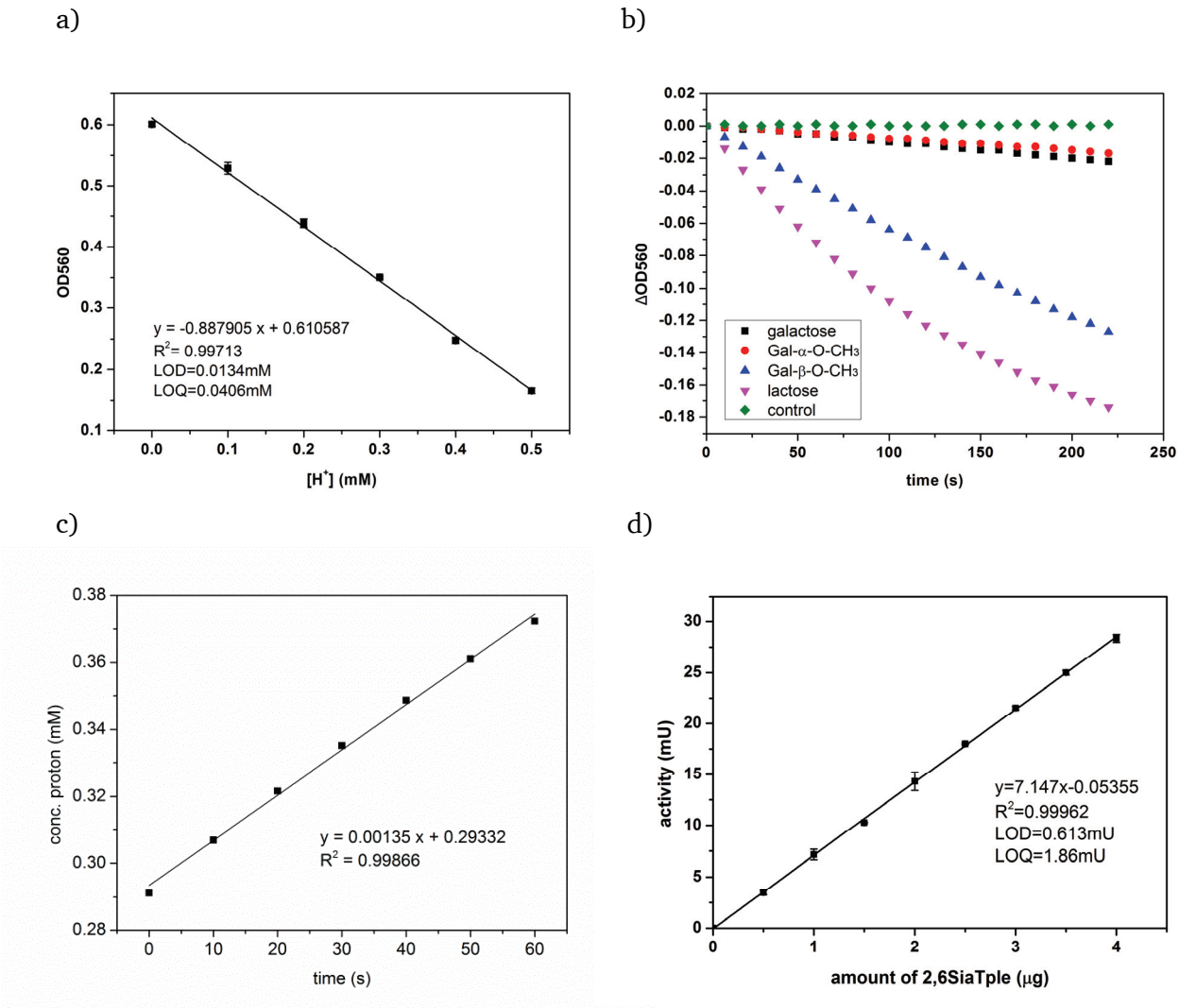



Figure 7.6. Development of the assay method. Samples contained 0.028 mM phenol red in 200 µL of 2 mM Tris buffer (pH 8.0) as well as enzyme and acceptor as specified. After addition of 0.5 mM CMP-Neu5Ac, absorbance was measured at 560 nm using a plate reader. a) Standard curve; reaction mixtures contained 2 µg 2,6SiaT$_{ple}$ and varying amounts of HCl (0–0.5 mM). b) Time curve of absorbance change using different acceptor substrates. Assay solutions contained 2 µg 2,6SiaT$_{ple}$ and various acceptors and control without enzyme. c) Time curve for absolute data. Absorbance data from (b) for lactose was converted into equivalent $H^+$ concentration according to the calibration curve. The slope of the resultant curve was used for calculating the specific activity. d) Dependence of rate on amount of enzyme; reaction mixtures contained lactose and varying amounts of 0 - 4 µg 2,6SiaT$_{ple}$.

### 7.2.5. Determination of kinetic constants of SiaT

The pH-based assay for SiaT is an exceptionally sensitive method, because the release of the proton affects the pH of the reaction system directly. Its continuous determination starting with the initial reaction rate makes itself to be very suitable for the measurement of the kinetic constants for SiaT with high definition. Using this method, the kinetic constants of both 2,3SiaT$_{pph}$ and 2,6SiaT$_{ple}$ were determined and listed in Table 7.1. Both the 2,3SiaT$_{pph}$ and 2,6SiaT$_{ple}$ have very similar $K_m$ and $k_{cat}$ values towards the donor CMP-Neu5Ac and the acceptor Lac. That means there is no significant difference between the behavior of 2,3SiaT$_{pph}$ or 2,6SiaT$_{ple}$ towards some natural substrates. However, 2,3SiaT$_{pph}$ and 2,6SiaT$_{ple}$ have totally different performance towards the galactosides with C1 modification in both α- and β-configurations,. 2,3SiaT$_{pph}$ binds methyl-α-D-galactopyranoside with a higher affinity methyl-β-D-galactopyranoside (the $K_M$ value of the latter is ca. 1.6 fold higher). The affinity of 2,3SiaT$_{pph}$ towards free Gal and Lac are very similar to that for methyl-β-D-galactopyranoside. On the other hand, the substrate selectivity of 2,6SiaT$_{ple}$ is ver different. While it shows practically identical affinity for the Lac disaccharide, the binding constant for Me-β-Gal is around 2.6 fold higher, which indicates that the second monosaccharide unit in the Lac acceptor has specific binding interaction with the active pocket of 2,6SiaT$_{ple}$. However, 2,6SiaT$_{ple}$ cannot accept α-galactosides as indicated by the fact that the $K_m$ value for methyl-α-D-galactopyranoside is too high to be measured. With α,β-anomeric, free Gal as acceptor, the $K_m$ value further increases to 234 mM, which is 12 fold higher than that for methyl-β-D-galactopyranoside, and 320 fold higher than that for Lac.

Table 7.1. Kinetic constants of 2,3SiaT$_{pph}$ and 2,6SiaT$_{ple}$ using the pH-based assay method

| | 2,3SiaT$_{pph}$ | | | 2,6SiaT$_{ple}$ | | |
|---|---|---|---|---|---|---|
| | $K_M$ (mM) | $k_{cat}$ ($s^{-1}$) | $k_{cat}/K_M$ ($s^{-1}\cdot mM^{-1}$) | $K_M$ (mM) | $k_{cat}$ ($s^{-1}$) | $k_{cat}/K_M$ ($s^{-1}\cdot mM^{-1}$) |
| CMP-Neu5Ac | 0.20±0.04 | 1.13±0.07 | 5.7 | 0.091±0.016 | 1.09±0.04 | 12 |
| Galactose (**1**)* | 0.72±0.07 | 1.41±0.02 | 2.0 | 234±40 | 1.11±0.13 | 0.0048 |
| methyl-α-D-galactopyranoside (**2**) | 0.39±0.02 | 1.51±0.01 | 3.9 | n.d.** | n.d. | n.d. |
| methyl-β-D-galactopyranoside (**3**) | 0.63±0.04 | 1.49±0.02 | 2.4 | 19.1±1.3 | 1.41±0.04 | 0.074 |
| Lactose (**11**) | 0.70±0.05 | 1.37±0.02 | 2.0 | 0.73±0.09 | 1.36±0.02 | 1.9 |

* See structures in Figure 7.8

**n.d. = not detectable

The kinetic constants of 2,3SiaT$_{pph}$ and 2,6SiaT$_{ple}$ at pH 6.0 and pH 8.0, respectively, have been reported by using a radioactive detection method with $^{14}$C labeled CMP-Neu5Ac as donor.[160, 161] The reported $K_m$ value for CMP-Neu5Ac is 0.05 mM for 2,3SiaT$_{pph}$ and 0.2 mM for 2,6SiaT$_{ple}$ which is similar to the value of 0.2 mM and 0.09 mM determined here. For the acceptors, the reported $K_m$ values of 2,3SiaT$_{pph}$ for Lac (Figure 7.8, **11**), methyl-α-D-galactopyranoside (Figure 7.8, **2**) and methyl-β-D-galactopyranoside (Figure 7.8, **3**) are 1.7 mM, 0.54 mM and 1.3 mM, respectively, which are all around 2 fold higher than those determined by the continuous pH-based assay, although the relative proportion between these data is consistent. The reported $K_m$ value of 2,6SiaT$_{ple}$ for Lac and methyl-β-D-galactopyranoside are 12.7 mM and 14.8 mM, respectively. The former is very different compared to the data (0.73 mM) obtained here, and is about 16 fold higher, while the latter value is similar to that (19 mM) determined here. Since the pH-based assay is the only continuous method of the three known assay methods and the kinetic constants were measured at different reaction conditions, the data discrepancy is plausible. The difference between the two $K_m$ values of 2,SiaT$_{ple}$ for Lac is huge. However, the corresponding apparent $K_m$ value of 2,6SiaT from *Photobacterium* sp. JT-ISH-224 (2,6SiaT$_{psp}$) is 31 mM for Lac and 99 mM for methyl-β-D-galactopyranoside; and the $K_m$ value of 2,6SiaT$_{pda}$ is 6.82 mM for Lac and 174 mM for methyl-β-D-galactopyranoside (Table 7.2). Both of these two homologous 2,6SiaTs have obvious $K_m$ value difference between that for Lac and for methyl-β-D-galactopyranoside. Therefore, the $K_m$ value of 2,6SiaT$_{ple}$ determined here for Lac and methyl-β-D-galactopyranoside is reasonable. From the structure analysis, the unknown three-dimensional structure of 2,6SiaT$_{ple}$ can be simulated computationally based on another 2,6SiaT from *Photobacterium* sp. JT-ISH-224 (2,6SiaT$_{psp}$) as a template, which has high homology to 2,6SiaT$_{ple}$ and for which an X-ray protein crystal structure has been determined.[148, 149, 210] The acceptor binding position could be identified by structure alignment (Figure 7.7). Asn157 is the corresponding amino acid residue for binding the OH7 of Lac, which could contribute to the orientation of Lac in a correct position. However, methyl-β-D-galactopyranoside has no binding contribution from such a Glc moiety, which could give methyl-β-D-galactopyranoside more flexibility in the active pocket of 2,6SiaT$_{ple}$.

| Table 7.2. Comparison of $K_M$ values (mM) of 2,3SiaT$_{pph}$ and 2,6SiaT$_{ple}$ with the reported data | | | | | | |
|---|---|---|---|---|---|---|
| | 2,3SiaT$_{pph}$ | | 2,6SiaT$_{ple}$ | | 2,6SiaT$_{psp}$ | 2,6SiaT$_{pda}$ |
| | pH assay | adioactive method[161] | pH assay | adioactive method[160] | adioactive method[162] | adioactive method[158] |
| CMP-Neu5Ac | 0.20 | 0.05 | 0.091 | 0.2 | - | 0.32 |
| Galactose | 0.72 | - | 234 | - | - | - |
| methyl-α-D-galactopyranoside | 0.39 | 0.54 | n.d.* | - | - | - |
| methyl-β-D-galactopyranoside | 0.63 | 1.3 | 19.1 | 14.8 | 99 | 174 |
| Lactose | 0.70 | 1.7 | 0.73 | 12.7 | 31 | 6.82 |
| **n.d. = not detectable | | | | | | |

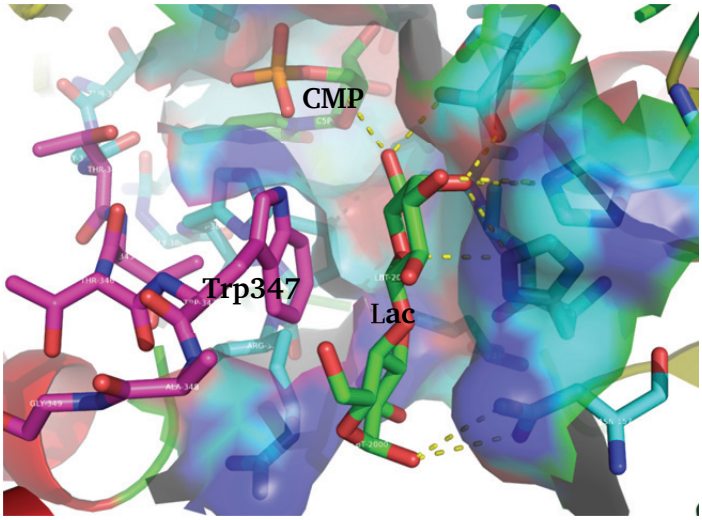



Figure 7.7. Lactose acceptor binding position simulated for 2,6SiaT$_{ple}$ by structure alighment with the X-ray structure of 2,6SiaT$_{psp}$ (PDB code 2Z4T).

### 7.2.6. Sialyltransferase screening for substrate tolerance

Besides the determination of kinetic constants, the pH-based assay method can also be used to assess the substrate tolerance of SiaT. A group of 13 potential and known acceptors were chosen for the specificity determination, which includes α- and β-galactosides, configurationally and structurally related monosaccharide and their derivatives, as well as lactose and lactosides, and the compound **12** and **13** for the further fluorescent labeling by a "clickable" reaction (Figure 7.8). Apparently, 2,3SiaT$_{pph}$ has quite broad acceptor range,

showing relatively high activity of 6-10 U/mg towards almost all 13 acceptors except for GlcNAc (around 2 U/mg), which comprises both configurational change at C4 and substitution at C2 of the natural Gal acceptor structure. Other monosaccharide than galactosides, for example Man, Glc and GalNAc, are well accepted by 2,3SiaT$_{pph}$ with only small decrease in activity as compared to Gal. This result illustrated that 2,3SiaT$_{pph}$ has an astonishing tolerance towards modifications at the C2 and C4 positions of the monosaccharide acceptors. On the other hand, 2,6SiaT$_{ple}$ presents very different behavior when comparing to 2,3SiaT$_{pph}$. Its activity towards α-galactosides, Glc, Man, GalNAc and GlcNAc are only 5% or less of that towards β-galactosides and lactosides. Its relative small acceptor range reveals that the configuration of hydroxyl groups and the anomeric linkage at C1, C2 and C4 highly important for the interaction of the enzyme with the acceptor substrate. Additionally, the modification of C1 of Lac does not reduce the acceptance of 2,3SiaT$_{pph}$ and 2,6SiaT$_{ple}$. As a result, the fluorophor-labeling and possibly surface immobilization of Lac acceptor will likely not affect the activity of SiaT.

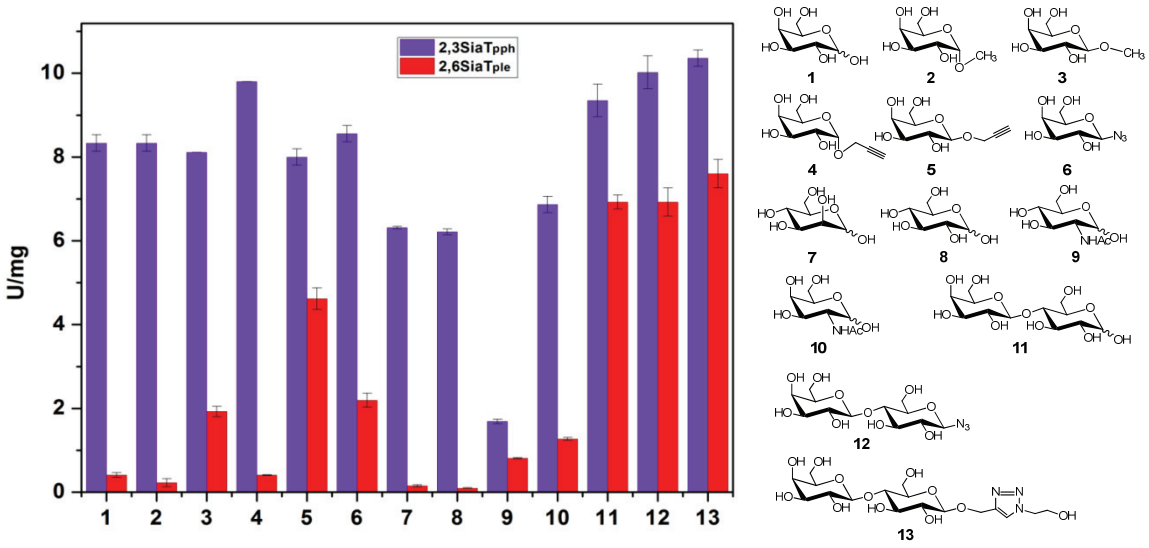



Figure 7.8. Substrate tolerance screening of 2,3SiaT$_{pph}$ and 2,6SiaT$_{ple}$. Reaction contained 0.028 mM phenol red in 200 µL of 2 mM Tris buffer (pH 8.0) as well as enzyme and 20 mM acceptor as specified. After addition of 1 mM CMP-Neu5Ac, absorbance was measured at 560 nm using a plate reader.

## 7.3. Conclusion

SiaT is an essential catalyst for the regiospecific chemoenzymatic synthesis of neo-sialoconjugates *in vitro*. 2,3SiaT$_{pph}$ and 2,6SiaT$_{ple}$ have been cloned and expressed in the *E. coli* expression system. Both enzymes could be easily purified in soluble form by Ni-affinity chromatography with high purity and activity. In order to understand the substrate specificity of these two SiaTs, a pH-based assay method has been established for determination of their kinetic constants and substrate tolerance, using phenol red as pH indicator in a low buffer system, and. This pH assay method is rapid, highly sensitive, and can be operated in continuous mode. With this method, the kinetic constants and the acceptor substrate tolerance of 2,3SiaT$_{pph}$ and 2,6SiaT$_{ple}$ has been determined. These two SiaTs have quite different acceptor specificity. While the 2,3SiaT$_{pph}$ can well accept both α- and β-galactosides, lactosides, Glc, Man and GalNAc, in striking contrast, the 2,6SiaT$_{ple}$ has high activity only towards β-galactosides and lactosides.

# 8. Directed Evolution of Cytosine-5′-Monophosphate- *N*-Acetylneuraminate Synthetase from *Neisseria meningitides* for Improved Tolerance towards 5′-Modified *N*-Acetylneuraminate analogues

## 8.1. Introduction

CMP-sialic acid synthetase (CSS, E.C. 2.7.7.43) catalyzes the CMP activation of sialic acid using CTP as donor and $Mg^{2+}$ as cofactor,[138] which is a key step of the Leloir biosynthetic route of complex oligosaccharides followed by a sialyltransfer reaction to form the specific 2,3-, 2,6- or 2,8-sialated products.[211] Besides its occurrence in higher vertebrates, CSS can also be found in several microorganisms, notably human pathogens, such as *Clostridium thermocellum*, *Streptococcus agalactiae*, *Escherichia coli*, *Haemophilus ducreyi*, *Mannheimia haemolytica, Pasteurella hemolytica* and *Neisseria meningitides*,[139] in which it is involved in the synthesis of host-like oligosaccharides to protect themselves against an attack by the host immune system.[212, 213] The CSS from *Neisseria meningitidis* ($CSS_{nme}$) is a well-studied enzyme that has been cloned[141] and its crystal structure analysed.[142] Recently, kinetic constants for several substrates analogs have been well determined by a new pH-based assay method and used to explain by acceptor binding with reference to the CSS from *Mus musculus* ($CSS_{mmu}$).[143] $CSS_{nme}$ has a remarkably broad acceptor tolerance towards 5', 7', 8' and 9' modified sialic acids.[143, 210] A series of Neu5Ac analogues carrying modifications of the acylamino group could be accepted by $CSS_{nme}$, although their corresponding $k_{cat}/K_M$ values were found to be much lower than those for Neu5Ac.[143] In order to improve its activity towards 5'-modified Neu5Ac analogues by site-directed mutagenesis, in this chapter an improved realignment of available three-dimensional CSS protein structures of $CSS_{mmu}$ and $CSS_{nme}$ was attempted with interpretation for the dynamic motions within the active site.

## 8.2. Results and discussion

### 8.2.1. Structure analysis of CSS

The crystal structure of CSS from *Neisseria meningitidis* ($CSS_{nme}$) has been determined with CDP as co-crystallized substrate in 2 conformations.[142] Its acceptor domain structure has been revealed by docking Neu5Ac to the active pocket with AutoDock,[142] which was a rational computationally simulated result based on the opened conformation of $CSS_{nme}$ towards the acceptor. However, the substrate access is a process of subtle interactions between both substrate and peptide, during which the positions of the corresponding residues or even the shape of the peptide might be changing dynamically. The closed conformation of the $CSS_{mmu}$ active site at least has slight changes in comparison with the opened one from $CSS_{nme}$. Therefore, the docking to an opened conformation in $CSS_{nme}$ might give biased structure information of the real Neu5Ac binding position. Up to date, only the CMP-Neu5Ac binding position in another CSS from *Mus musculus* ($CSS_{mmu}$) has been revealed by crystal structure determination of its product-liganded form.[214, 215] $CSS_{mmu}$ has 24.6% identity and 40.1% consensus homology with $CSS_{nme}$ (Sequence 8.1). Therefore, more detailed information on substrate binding interactions in the active site of $CSS_{nme}$ might be devired by a careful analysis of the three-dimensional structure alignment between $CSS_{nme}$ and $CSS_{mmu}$.

```
                1                                                          75
CSSnme     (1)  MEKQNIAVILARQNSKGLPLKNLRKMNGISLLGHTINAAISSKCFDRIIVSTDGGLIAEEAKNFGVEVVLRPAEL
CSSmmu     (1)  -PPHLAALVLARGGSKGIPLKNIKRLAGVPLIGWVLRAALDAGVFQSVWVSTDHDEIENVAKQFGAQVHRRSSET

               76                                                         150
CSSnme    (76)  ASDTASSISGVIHALETIGSNSGTVTLLQPTSPLRTGAHIREAFSLFDEKIKGSVVSACPMEHHPLKTLLQINNG
CSSmmu    (75)  SKDSSTSLDAIVEFLNYHN-EVDIVGNIQATSPCLHPTDLQKVAEMIREEGYDSVFSVVRRHQFRWSEIQKGVR-

               151                                                        225
CSSnme   (151)  EYAPMRHLSDLEQPRQQLPQAFRPNGAIYINDTASLIANNCFFIAPTKLYIMSHQDSIDIDTELDLQQAENILNH
CSSmmu   (148)  EVTEPLNLNPAKRPRRQDWDGELYENGSFYFAKRHLIEMGYLQGGKMAYYEMRAEHSVDIDVDIDWPIAEQRVLR

               226
CSSnme   (226)  KES----
CSSmmu   (223)  FGYFGKE
```

Sequence 8.1. Protein sequence alignment of $CSS_{nme}$ and $CSS_{mmu}$

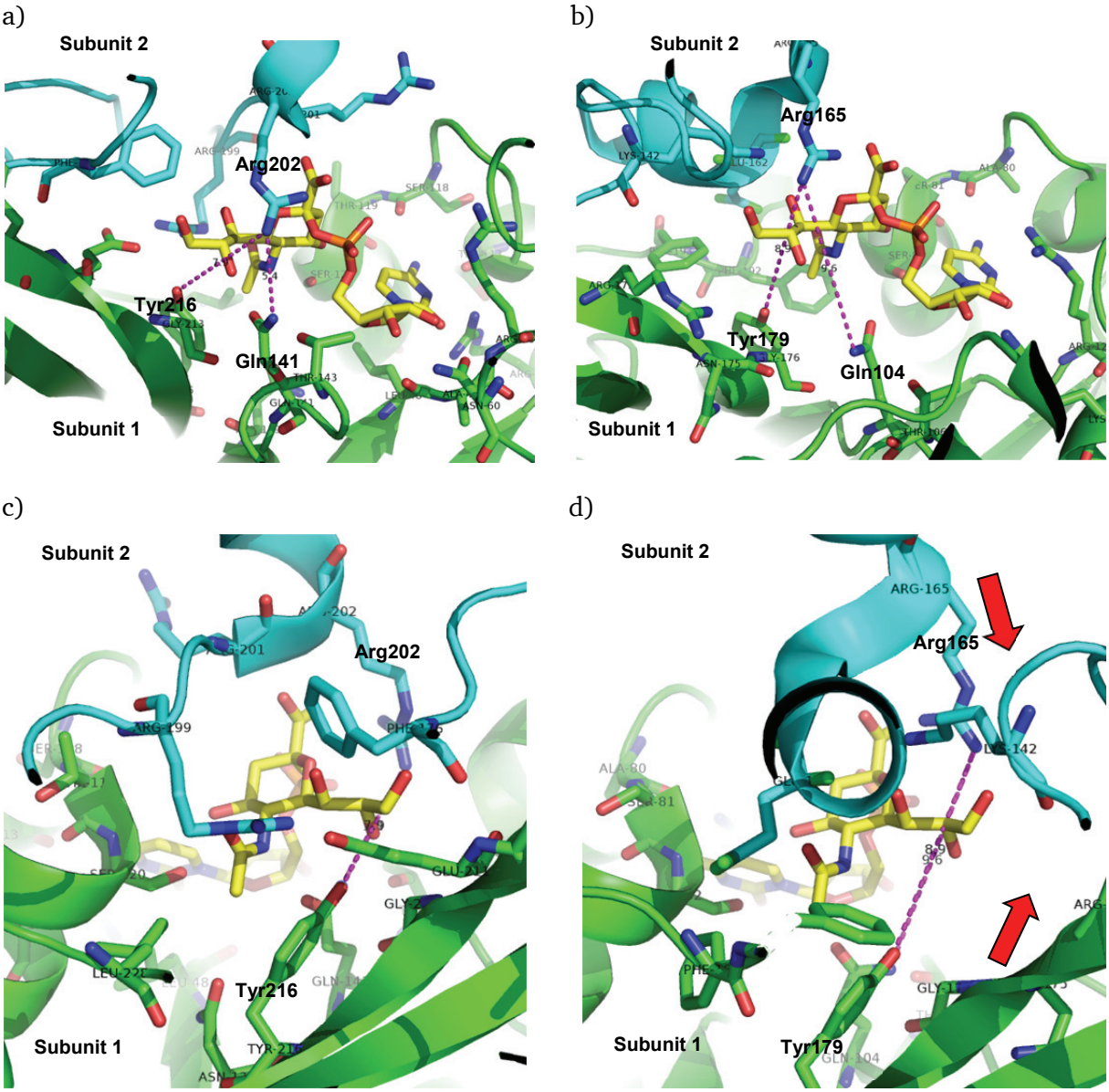


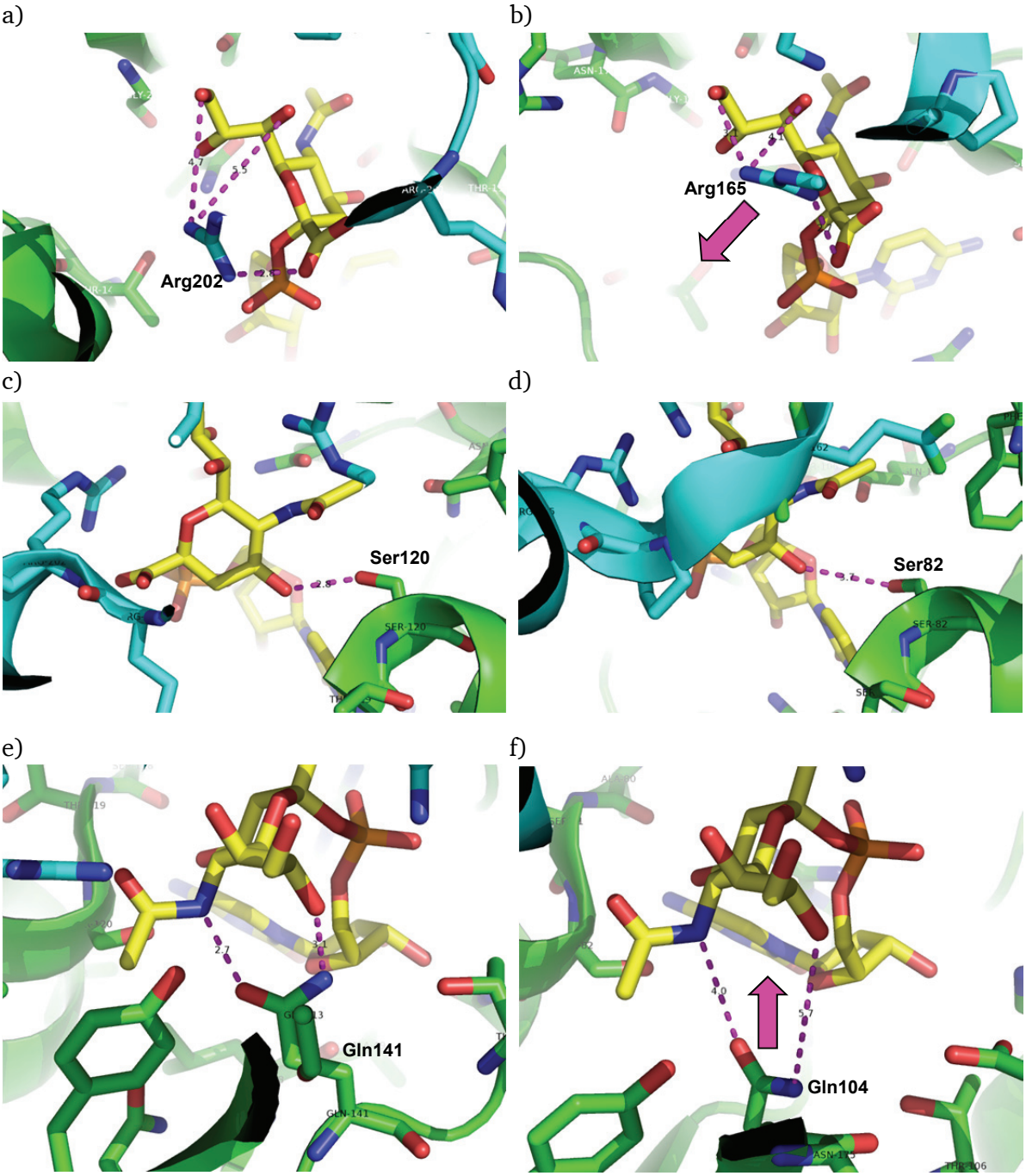

Figure 8.1. Overall structure of the active pocket of $CSS_{nme}$ (a and c) in comparison to that of $CSS_{mmu}$ (b and d). The active residues identified in $CSS_{mmu}$ Tyr216, Gln141 (in subunit 1, green) and Arg202 (in subunit 2, blue) are exactly corresponding to Tyr179, Gln104 (in subunit 1) and Arg165 (in subunit 2) in $CSS_{nme}$. The distance measurement is marked as purple dotted lines. The motion trends of the secondary structures are marked as red arrow.

Although $CSS_{nme}$ and $CSS_{mmu}$ are two heterologous isozymes, the three-dimensional structure alignment of their acceptor binding pockets showed that they are provided with a very similar catalytic site, which indicates they likely share the same catalytic mechanism. In both the dimeric enzymes, the active site is jointly constructed by both subunits.[141, 142] The active residues identified in $CSS_{mmu}$ Tyr216, Gln141 (in subunit 1) and Arg202 (in subunit 2) are exactly corresponding to Tyr179, Gln104 (in subunit 1) and Arg165 (in subunit 2) in $CSS_{nme}$. (Figure 8.1a and 8.1b) Therefore, measuring the distances between these identical residues can be used to illustrate the structure difference between opened and closed conformation of

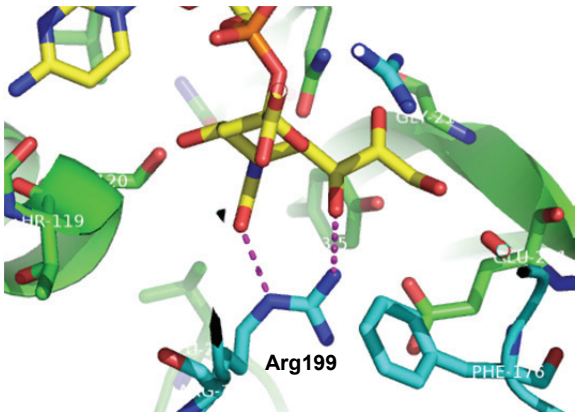
$CSS_{nme}$ and assist in the identification of the Neu5Ac binding position in $CSS_{nme}$. As shown in Figure 8.1a, the distances between Arg202 to Tyr216 and Gln141 are 7.9Å and 5.4Å, respectively. However, corresponding distances in $CSS_{nme}$ are significantly enlarged 8.9Å and 9.5Å, respectively. (Figure 8.1b) This comparison demonstrates that the crystal structure of $CSS_{nme}$ is indeed in an opened conformation to facilitate acceptor substrate access. After capture of Neu5Ac, owing to the interaction between the substrates and amino acid residues, the two subunits of $CSS_{nme}$ will perform a tighter interaction. The β-sheet 1 and the α-helix 1 have to move jointly towards each other to close in on the Neu5Ac substrate (Figure 8.1d, red arrows). The side chains of Tyr179, Gln104 and Arg165 could form the arms surrounding the acceptor almost totally. (Figure 8.1c and 8.1d) These peptide movements will result in an adjustment of the position of all active residues for binding of the acceptor, which is indispensable for any detailed analysis of active residues.

From the model alignment of CMP-Neu5Ac (ligand conformation from $CSS_{mmu}$) in the $CSS_{nme}$ active pocket, the amino group of Arg165 would have 2.7 Å, 3.1 Å and 4.1 Å closest distances to $COO^-$, 9-O and 7-O of the Neu5Ac moiety, respectively (Figure 8.2b). It seems to have the broad possibility to form hydrogen bonds with all of these groups.[143] However, when the active pocket is closed, Arg165 will more towards the β-sheet 1 from the other subunit as seen from the location of the corresponding Arg202 in $CSS_{mmu}$ (Figure 8.2a). Since Arg165 is the essential[215] and the only basic residue nearby, it will keep its hydrogen bond with $COO^-$, but will certainly dislocate from 9-O and 7-O, keeping about a 5 Å distance to those as indicated by the situation of corresponding Asp202 in $CSS_{mmu}$ (Figure 8.2a). Therefore, Arg165 may mostly form an ionic bond to $COO^-$ instead of hydrogen bouding to 9-O and 7-O. For 4-O, Ser82 is the only residue at a satisfactory distance (3.7 Å) and orientation to form a hydrogen bond (Figure 8.2d). Gln can act both as an acidic or basic group because of its amide moiety. In $CSS_{mmu}$, Gln141 binds with both 5-NH and 8-O perfectly with the distance of 2.7 Å and 3.1 Å (Figure 8.2e). The corresponding residue Gln104 in the opened $CSS_{nme}$ conformation is somewhat farther away from 5-NH and 8-O. (Figure 8.2f) However, as described above, Gln104 has a clear trend towards Neu5Ac during the acceptor binding. (Figure 5.2d, red arrow) Therefore, Gln104 is still the most plausible functional residue for binding of both 5-NH and 8-O in $CSS_{nme}$. Lys142 in $CSS_{nme}$ has only a 2.7 Å distance to 7-O. Since it locates on a rather flexible loop, it can be identified as the corresponding residue to binding of 7-O (Figure 8.2h). For 9-O, there is no suggested binding contact in $CSS_{mmu}$[142] (Figure 8.2i). However, the kinetic study of $CSS_{nme}$ has shown that 9-O has a positive binding contribution.[143] Indeed, when β-sheet 1 approaches the Neu5Ac moiety, the extensible side chain of Arg173 can reach out to 9-O in a reasonable distance from 5.9 Å to an approximated

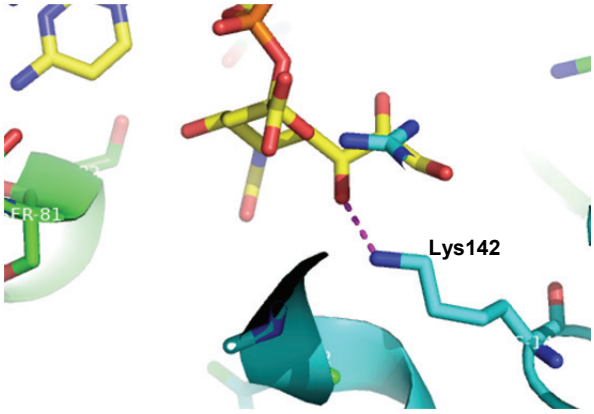
of around 3-4 Å (Figure 8.2j). In $CSS_{mmu}$, the binding residue to the C=O group of the *N*-acyl moiety is Arg199 (Figure 8.2k). However, only Glu162 is located at the corresponding position in the subunit 2 of $CSS_{nme}$ with a distance of 5.8 Å to the carbonyl group (Figure 8.2l). When the α-helix 1 changes its position towards the closed conformation, the distance between Glu162 and =O will also be shortened. Although the side chain of Glu is shorter than that of Arg, it could still reach the effective distance to C=O. Thus, it should be the corresponding binding contact for the *N*-acyl group. For binding of the non-polar $CH_3$ of the amide moiety, the aromatic aide chain of Phe192, Phe193 and Tyr179 could jointly form a hydrophobic pocket[143] (Figure 8.2n).



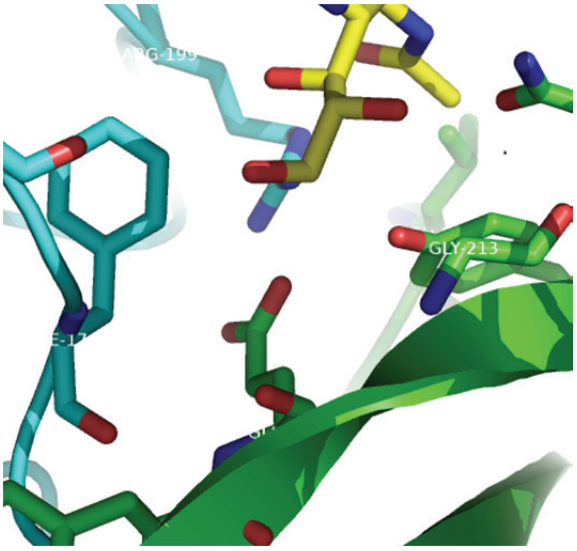
g)


Arg199

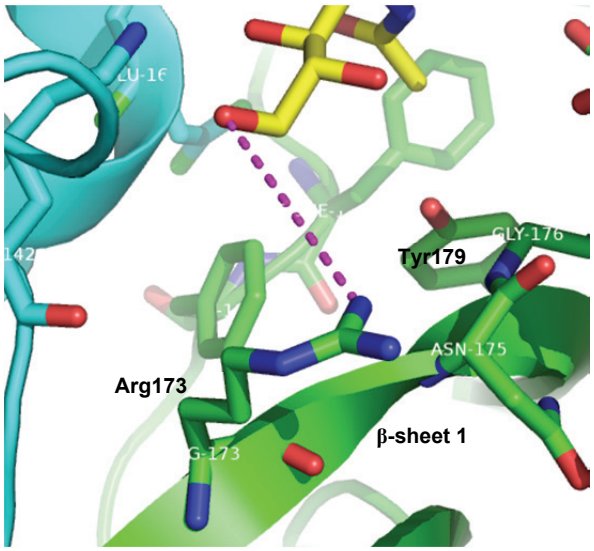
h)


Lys142

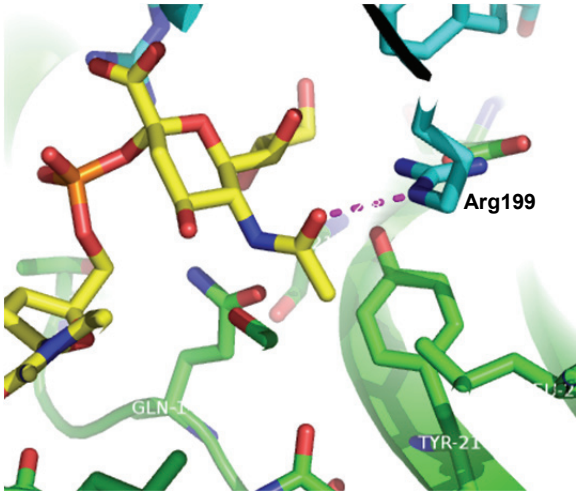
i)


GLY-213

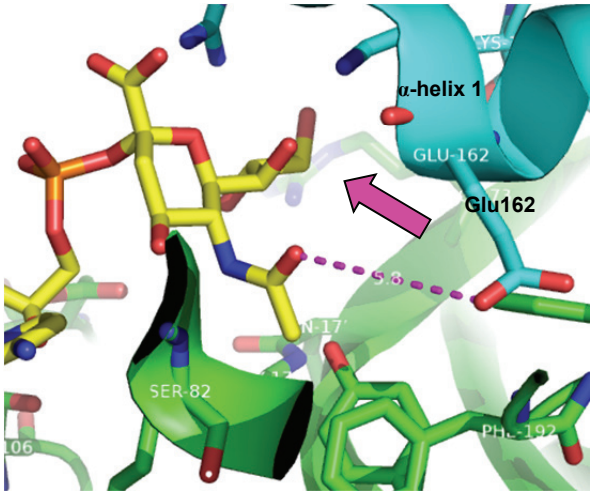
j)


Tyr179
GLY-176
ASN-175
Arg173
β-sheet 1

k)


Arg199
TYR-21

l)


α-helix 1
GLU-162
Glu162
SER-82

m)

n)

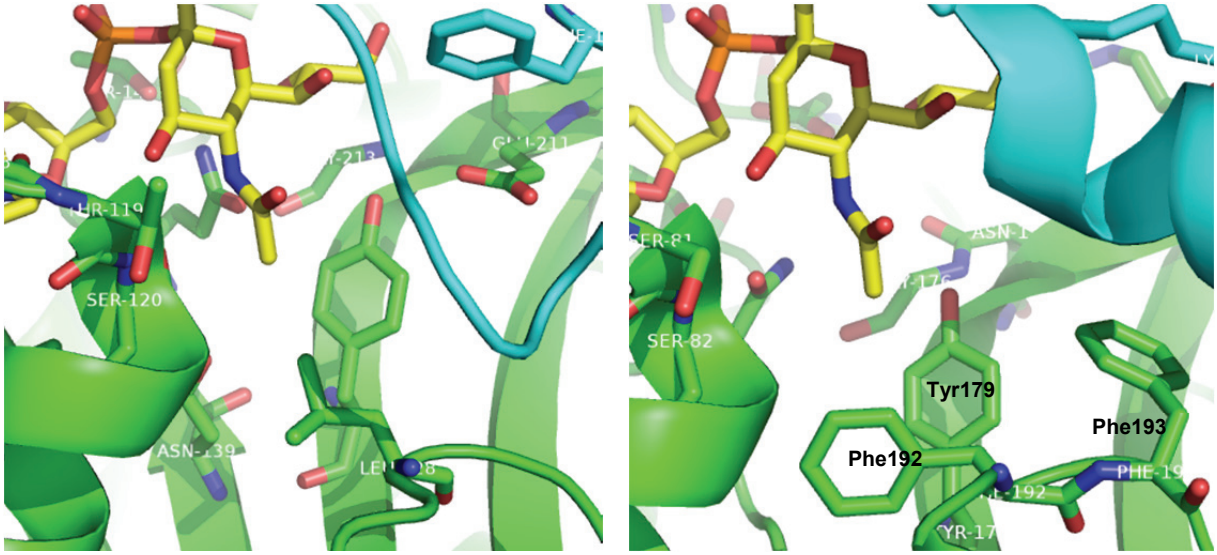



Figure 8.2. Active pocket analysis from structure alignment. Left: $CSS_{mmu}$, right: $CSS_{nme}$. Distance measurements are marked by purple dotted lines. The motion trends of the secondary structures are marked by purple arrows. Green: subunit 1. Blue: subunit 2. (See description in details in 8.2.1)

When compared to the substrate docking results,[142] the suggested residues in contact with $COO^-$, 4-O, 5-NH, 7-O and 8-O are identical. (Table 8.1) Two more residues, Asp209 and Thr106, were also considered from the docking result for 8-O. However, Asp209 has been involved in the co-binding to $Mg^{2+}$ together with Asp211.[216] Thr106 has an estimated 5.2 Å distance to 8-O in the open conformation of $CSS_{nme}$. It could get nearer to 8-O during the closing movement, though there remains a 4.3 Å distance in $CSS_{mmu}$. As a result, Gln104 most likely forms the contact to 8-O. Try179 was suggested to have a hydrogen bonding interaction with 9-O.[142] According to the X-ray structure of the $CSS_{mmu}$, the corresponding Tyr216 has has a large 5.0 Å distance to 9-O. The only possibility for a binding interaction would be that the carbon-oxygen bond of 9-O rotates to a *cis* position with 8-O (Figure 8.2j), which, however, is not its energy minimium orientation. Thus, Tyr179 in $CSS_{nme}$ might be less probable for making a hydrogen bond interaction with 9-O. Compared to another structure alignment result,[143] Arg165 probably has no immediate contacts to 7-O and 9-O because of its shifted position described above (Figure 8.2a and b). Asn175, Gly176 and Tyr179 are located on the same rotating β-sheet. The side chains of Asn175 and Gly176 reache out to the other side against Neu5Ac (Figure 8.3). Thus, they are not at the appropriate location to contacting 8-O. Try179 has indeed a trend to move towards 8-O, but owing to its rigid location on a β-sheet, it might not have enough flexiblility to shorten the distance from 5.2 Å to less than 4 Å (Figure 8.3). Therefore, in comparison its corresponding Tyr216 in $CSS_{mmu}$ still keep 4.3 Å away from 8-O.

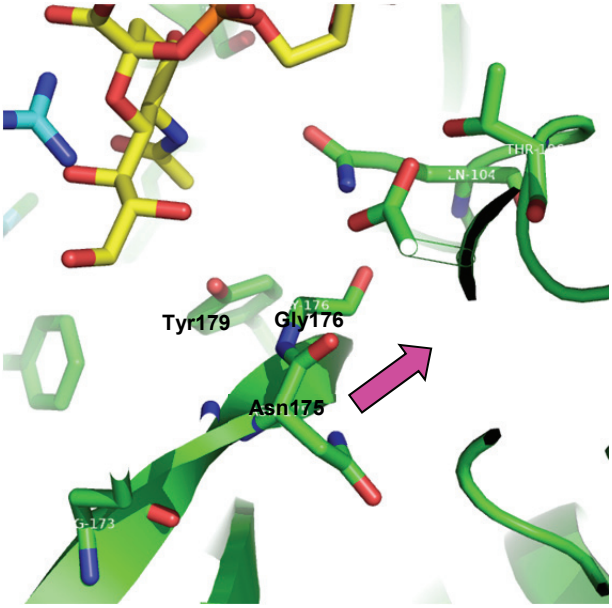



Figure 8.3. Location of Asn175, Gly176 and Tyr179 in the active site of the open form of $CSS_{nme}$. The motion trends of the secondary structures are marked by purple arrows. Green: subunit 1. Blue: subunit 2.

Table 8.1. Specific interactions suggested between residues of *N. meningitides* CSS to CMP-Neu5Ac

| Atom or group in Neu5Ac | Contacting residue in murine CSS[214, 215] | Contacting residue suggested by docking[142] | Potential additional or alternative contacts from structure alignment | |
|---|---|---|---|---|
| | | | ref.[171] | in this work |
| $COO^-$ | Arg202 | Arg165 | | Arg165 |
| 4-O | | Ser82 | | Ser82 |
| 5-NH | Gln141 | Gln104 | | Gln104 |
| 7-O | | Lys142 | Arg165 | Lys142 |
| 8-O | Gln141 | Gln104, Asp209, Thr106 | Asn175, Gly176, Tyr179 | Gln104 |
| 9-O | | Tyr179 | Arg165, Arg173 | Arg173 |
| =O (*N*-acyl) | Arg199 | | | Gln162 |
| $CH_3$ | Ile124, Tyr216, Leu228 | Phe192, Phe193 | Tyr179 | Phe192, Phe193, Tyr179 |

In summary, according to the structure alignment results described above, it is suggested that Arg165, Ser82, Gln104, Lys142, Gln104, Arg173 and Gln162 are the amino acid residues most likely offering hydrogen bonding interaction to $COO^-$, 4-O, 5-NH, 7-O, 8-O, 9-O and C=O (*N*-acyl) of the Neu5Ac moiety, respectively. Phe192, Phe193 and Tyr179 potentially form a hydrophobic pocket to accept the non-polar $CH_3$ group (Table 8.1).

### 8.2.2. Building and screening of single site mutagenesis libraries for Phe192 and Phe193

From the kinetic constant study of $CSS_{nme}$, 5'-modified Neu5Ac analogues are not well accepted as indicated by 10-100 folds increased $K_M$ values and 30-1700 fold decreased $k_{cat}/K_M$ values in comparison with natural substrate Neu5Ac.[143] According to the structure alignment of $CSS_{nme}$, Phe192 and Phe193 are directly contacting the acylamino moiety of Neu5Ac (Figure 8.4). Apparently, the phenyl groups of Phe192 and Phe193 block the access of modified groups to the pocket. A study of mutants from an Ala scan of $CSS_{nme}$ had shown that Phe192Ala causes a remarkable activity improvement towards Neu5Ac-OBz. Therefore, in order to improve the acceptance of $CSS_{nme}$ towards 5'-modified Neu5Ac analogs, focussed mutagenesis libraries were planned targeting on these two positions to obtain improved mutants offering less steric hindrance or a more polar pocket. Two individual single site libraries at positions Phe192 and Phe193 were constructed firstly to reveal their independent effect against modified groups, to be followed by a combined double site library for both Phe192 and Phe193 to illustrate their cooperative action.

For the single-site libraries, the gene codons on the mutation site were designed as NNK covering the entire set of all 20 amino acids for reducing the library size to a lower limit of at least 95 colonies, while still guaranteeing a statistical library coverage of 95%,[26] which could easily be performed on a single 96-well plate. The QuikChange method was used for the construction of the mutagenesis libraries, which has been used widely in saturation mutagenesis. The common protocol uses a PCR to produce mutant plasmids followed by *Dpn*I digestion, transformation and isolation of positive mutants on selection plates. A total of 96 colonies were picked randomly for each single site library (named Lib-F192k and Lib-F193k, respectively).

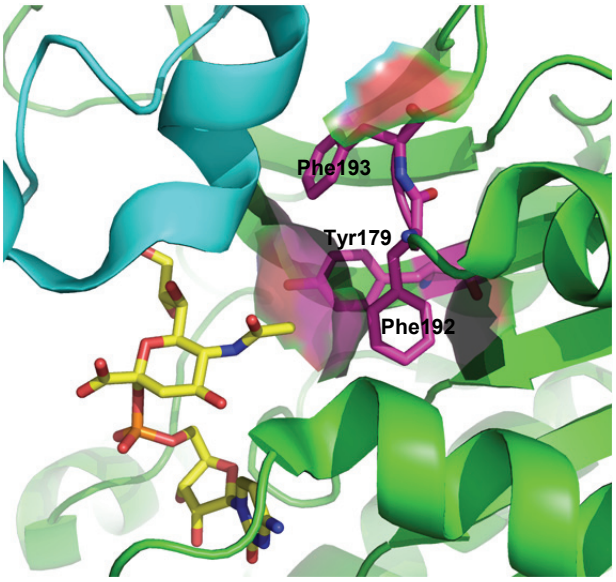



Figure 8.4. Mutation sites of $CSS_{nme}$. Phe192 and Phe193 are directly contacting the acylamino moiety of Neu5Ac. Apparently, the phenyl groups of Phe192 and Phe193 block the access of modified groups to the pocket.

### 8.2.3. Building and screening of a Phe192-Phe193 double site mutagenesis library

For building the double site library (Phe192 and Phe193), there were two well-used strategic alternatives. One is named Iterative Saturation Mutagenesis (ISM),[217] in which the most active single-site mutants at either position (Phe192 or Phe193) are identified first, followed by performing a second single site mutagenesis on the other Phe site using these positive mutants as templates. This is a relative simple and efficient method with the advantage of small library size, especially when only a few mutant sites need to be induced. However, a few potentially interesting mutants might unfortunately be missed with this method, because the first (or more) mutant site is fixed early on in these libraries. This will limit the diversity of mutations, and limits the opportunity to find cooperative interactions between any two (or more) amino acid residues. The other strategy is to perform saturation mutagenesis on both positions simultaneously. When using the NNK codon on both Phe sites, at least 3,066 colonies should be included in this library for a statistical 95% coverage of the library for all 20 residues,[26] so that the total required library size becomes much larger than that of sequential single-site ones. However, if we use NRT or NDT degeneracy, the library size could be reduced to 190 or 430 colonies at 95% library coverage with a restriction to only 8 or 12 amino acid residues, respectively. However, this would still include the typical residues containing polar, nonpolar (unavailable with NRT degeneracy), charged and aromatic

building blocks in the library.[26] Thus, two double-site libraries were built with either NRT or NDT degeneracies to compare their forward mutation efficiency for our experimental subject and to obtain as many positive hits as possible. Because the mutation sites are adjacent, the QuikChange method was also used for the construction of the mutagenesis libraries. A total of 192 and 432 colonies were chosen randomly for two double-site libraries (named Lib-F1923r and Lib-F1923d), respectively.

### 8.2.4. Screening and docking of CSS mutants

For the screening of the CSS mutagenesis libraries, a high-throughput method is required. Previously, a pH-based assay method has been well established for CSS activity determination, in which cresol red was used as pH indicator in 2 mM Tris as a low concentration buffer system.[143] With this assay method, the kinetic constants and substrate tolerance of $CSS_{nme}$ have been successfully measured towards various sialic acids and their derivatives. Because of its high throughput, sensitivity and economy, this pH-based assay also promised to be a suitable screening method for evaluation of the CSS mutagenesis libraries. To avoid inducing additional buffer components, a culture and lysis method described in the literature was used, which performed well in another pH-base screening system.[208] The cells were cultured and induced in LB medium without any additional buffer. The harvested cells were lysed by addition of 1/10 (V/V) BugBuster solution containing lysozyme and Benzonase.[208] The supernatant of the lysate was used directly as enzyme sample for the activity screening. The OD change was read and recorded by a plate reader at 574 nm. After the first round of screening, 16, 24 or 32 positive samples in each single-site and double-site libraries, respectively, were chosen for the rescreening. Finally, a total of 40 most active mutants from these four libraries were picked for the genetic identification. The screening and sequencing results are listed in Table 8.2.

Table 8.2. Screening and sequencing results for mutagenesis libraries

| Library | Neu5Ac-OBz | | Neu5Hex | | $Neu5PenN_3$ | |
|---|---|---|---|---|---|---|
| | Mutant (192/193) | Relative activity | Mutant (192/193) | Relative activity | Mutant (192/193) | Relative activity |
| Phe192k | S/F | 12.8 | S/F | 4.6 | S/F | 4.0 |
| | A/F | 12.4 | C/F | 4.2 | T/F | 3.7 |
| | C/F | 12.2 | A/F | 3.9 | A/F | 3.4 |
| | N/F | 11.3 | N/F | 3.6 | N/F | 3.2 |
| | T/F | 10.9 | P/F | 3.2 | P/F | 2.9 |
| | P/F | 7.3 | T/F | 2.9 | C/F | 2.8 |
| Phe193k | F/Y | 3.9 | F/Y | 2.4 | F/Y | 2.1 |
| | F/I | 3.2 | F/I | 1.8 | F/I | 1.8 |
| Phe1923r | C/Y | 14.1 | C/Y | 3.9 | S/Y | 3.8 |
| | S/Y | 12.0 | S/Y | 3.5 | S/H | 2.9 |
| | S/H | 10.7 | S/H | 3.0 | C/Y | 2.8 |
| | P/F | 7.3 | | | | |
| Phe1923d | C/F | 13.3 | C/F | 4.6 | I/Y | 3.6 |
| | S/H | 10.3 | S/I | 3.5 | S/H | 3.0 |
| | C/Y | 10.0 | G/I | 3.2 | S/I | 3.0 |
| | S/I | 10.0 | G/Y | 3.0 | G/I | 2.8 |
| | S/V | 7.0 | S/H | 2.9 | C/F | 2.7 |
| | | | | | G/Y | 2.5 |

*The activity of wild-type is calculated as 1.

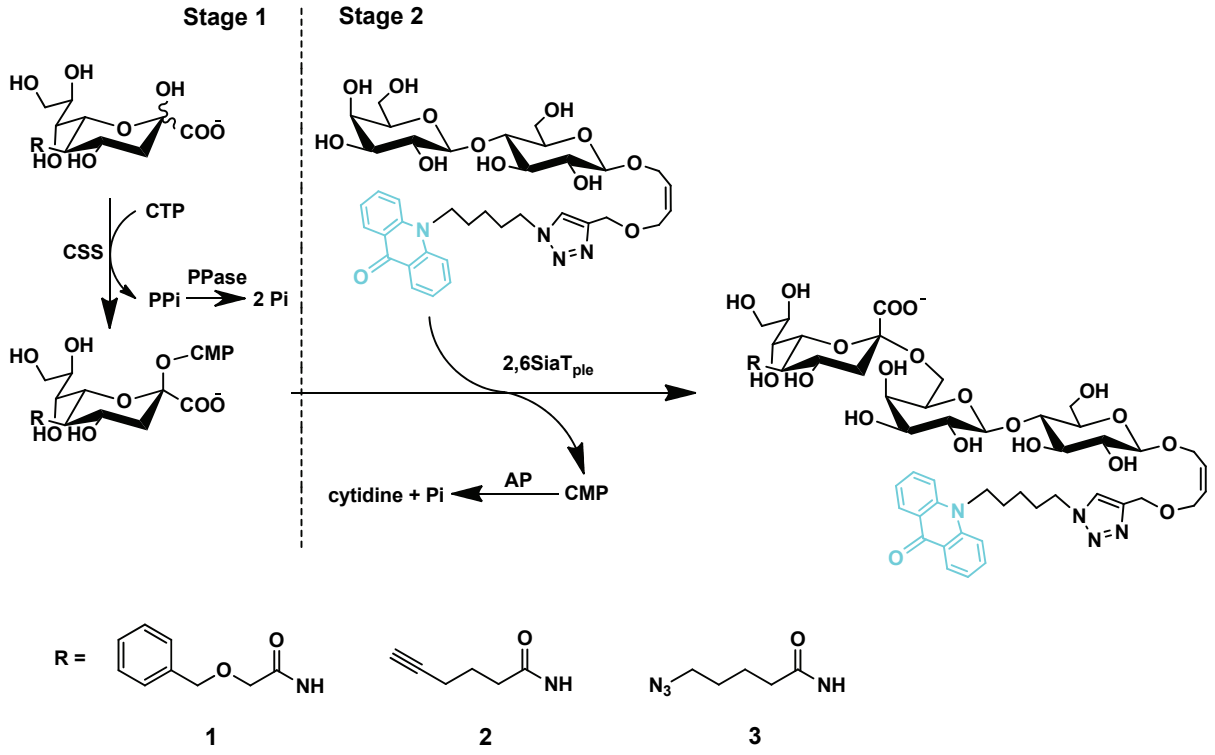
The single site mutation results for F192A and P193A revealed that the Phe192 site has more influence in binding of the substrate's acyl moiety than Phe193. This hypothesis is proved by the far better screening results of library Lib-F192k than those of Lib-F193k. Comparing the forward mutation efficiency between these two single-site libraries, much more and also higher active mutants can be found for Phe192, offering at least 2-3 fold higher activity than top mutants from Lib-F193k. The positive hits in Lib-F192k (Ser, Ala, Cys, Asn, Thr and Pro) are all those with amino acid residues that have shorter side chains and offer more space for the modified acyl groups than Phe. The top one F192S, which has an additional hydroxyl group, has a slightly higher activity than F192A. This particular hydroxyl group possibly forms an additional hydrogen bond to the modified acyl group of sialic acids, which is conducive to

the binding of substrate to enzyme. However, the most advanced mutants in the Lib-F193k library contained Tyr and Ile instead of Ala. Tyr has one more hydroxyl on the benzene ring than Phe, while Ile is a hydrophobic amino acid with a C4 side chian which is similar to but shorter than that of Phe, and much more voluminous than that of Ala. This result might reveal that the suitable side group volume and hydrophobicity or/and stable hydrogen bond at Phe193 are more important than an additional space and flexibility for the activity of CSS towards substrates modified at the acyl group.

From the double site libraries, more coevolution mutants were observed. The NDK degeneracy offers four more hydrophobic amino acid residues than the one based on NRK. Thus, the positive mutants in Lib-F1923d present more diversity than those in Lib-F1923r. On the Phe192 position, we obtained Cys, Ser, Pro and also Gly and Ile, which can fit the requirement of more room for the modified acyl group. Meanwhile, on the Phe193 site, His and Val were newly found, the side-chains of which are similar to Tyr and Ile. Although there are more variations obtained in the double site libraries, there seems to be no major activity improvement as compared to the Lib-F192k hits. In consideration of the relative inaccuracy of measurements when using crude extract samples, the seven top mutants (Cys/Phe, Cys/Tyr, Ser/Phe, Ser/Tyr, Gly/Tyr, Ala/Phe and Ser/His) were chosen and purified by Ni-affinity chromatography, and their performance as pure enzymes will be compared in ongoing studies.

### 8.2.5. Verification by synthesis of neo-sialoconjugates

In order to validate the real improved activity towards 5' modified Neu5Ac, the *in situ* CMP activation of Neu5Ac analogues was applied by using three selected CSS variants, $CSS_{nme}$-F192C/F193Y, $CSS_{nme}$-F192S, and $CSS_{nme}$-F192S/F193Y, for the preparative synthesis of new neo-sialoconjugate products, in which the mutant CSS catalysis was coupled with the $2,6SiaT_{ple}$ conversion, which causes overall one-pot activation and transfer of the sialate moiety to Lac-D-T-P-Acr. (Scheme 8.1)




Scheme 8.1. Preparative synthesis of new neo-sialoconjugate products by using optimized mutants $CSS_{nme}$-F192C/F193Y, $CSS_{nme}$-PF192S, and $CSS_{nme}$-F192S/F193Y, in a one-pot, coupled system with 2,6$SiaT_{ple}$. 1: Neu5Ac-OBz. 2: Neu5Hex. 3: $Neu5PenN_3$.

From the kinetic study of $CSS_{nme}$, the wild-type has only low $k_{cat}/K_M$ values of 0.2 and 0.7 towards Neu5Ac-OBz and $Neu5PenN_3$, respectively.[143] Variants $CSS_{nme}$-F192C/F193Y, $CSS_{nme}$-F192S, and $CSS_{nme}$-F192S/F193Y were found as the top variants with 14.1 fold, 4.6 fold and 4.0 fold activity improvement towards Neu5Ac-OBz, Neu5Hex and $Neu5PenN_3$, respectively. During the synthesis of neo-sialoconjugate products, the activition step of Neu5Ac-OBz, Neu5Hex and $Neu5PenN_3$ could be completed in one hour. The coupled sialyltransfer reaction also confirmed that indeed the CMP-sialic acid derivatives were formed correctly because they were well accepted as donor by 2,6$SiaT_{ple}$. The yields of the final products were 50%, 64% and 70%, repectively.

## 8.3. Conclusion

$CSS_{nme}$ shows much less activity towards 5'-modified Neu5Ac analogues. From the active site alignment for the acceptor binding of $CSS_{nme}$, Phe192 and Phe193 have been identified as the corresponding residues for the acylamino group of Neu5Ac. Therefore, four saturation mutagenesis libraries were constructed at positions of Phe192 and Phe193 (single and combined mutating) to improve its activity towards three representative substrates Neu5Ac-OBz, Neu5Hex and $Neu5PenN_3$. Using a pH-based high-throughput screening method, a series of positive variants were ifentified by screening. Among them, $CSS_{nme}$-F192C/F193Y, $CSS_{nme}$-F192S, and $CSS_{nme}$-F192S/F193Y were the top candidates with 14.1-fold, 4.6-fold and 4.0-fold activity improvements towards Neu5Ac-OBz, Neu5Hex and $Neu5PenN_3$, respectively. In order to validate the real improved activity towards 5'-modified Neu5Ac, the *in situ* CMP activation of Neu5Ac analogues was applied by using these three CSS variants for a preparative synthesis of new neo-sialoconjugate products when coupled with $2,6SiaT_{ple}$ catalysis to obtain final product yields of 50%, 64% and 70%, respectively.

# 9. Directed Evolution of 2,6-Sialyltransferase from *Photobacterium leiognathi* JH-SHIZ-145 towards the Use of α-Galactosides as Glycosyl Acceptors

## 9.1. Introduction

2,6SiaT$_{ple}$ is a typical and efficient 2,6-sialyltransferase having an optimum pH at around 8.0.[160] In Chapter 4, the substrate tolerance and kinetic constants of 2,3SiaT$_{pph}$ and 2,6SiaT$_{ple}$ were evaluated based on a pH-based assay method and compared to literature results obtained by the HPLC methods.[160, 161] According to their specificity data, 2,3SiaT$_{pph}$ and 2,6SiaT$_{ple}$ showed different behavior towards α- and β-galactosides. Both α- and β-galactosides are well accepted by 2,3SiaT$_{pph}$, while only β-configurated galactosides are acceptable by 2,6SiaT$_{ple}$. Although these two enzymes belong to the same sialyltransferase family (GT80) in CAZy database (carbohydrate-active enzymes),[218] the structural difference between their active sites dominates their different substrate specificities. Fortunately, both the crystal structures of 2,3SiaT$_{pph}$ and another 2,6SiaT from *Photobacterium* sp. JT-ISH-224 (2,6SiaT$_{psp}$) similar to 2,6SiaT$_{ple}$ have already been determined.[148, 149, 210] Because of the high homology between 2,6SiaT$_{ple}$ and 2,6SiaT$_{psp}$ (53.8% identity and 67.8% consensus), the three-dimensional structure of 2,6SiaT$_{ple}$ can be well simulated computationally based on the 2,6SiaT$_{psp}$ structure as a template. The active domain comparison between 2,3SiaT$_{pph}$ and 2,6SiaT$_{ple}$ can provide the feasibility to reveal the structure difference between each other, to allow a possible site-directed mutagenesis towards an extension of substrate tolerance. As an 2,6-sialyltransferase, 2,6SiaT$_{ple}$ has shown an efficient catalytic specificity towards lactose derivatives.[171] For broader preparative-scale applications, it would be desirable if the enzyme could also accept α-galactosides as substrates. Thus, in this chapter, three-dimensional structures of 2,3SiaT$_{pph}$ and 2,6SiaT$_{ple}$ are aligned, especially for their active-site pockets. Based on this structural analysis, three site mutagenesis libraries of 2,6SiaT$_{ple}$ were built in an attempt improving its activity towards α-galactosides.

## 9.2. Results and discussion

### 9.2.1. Active domain analysis of 2,3SiaT$_{pph}$ and 2,6SiaT$_{ple}$

The crystal structure of 2,3SiaT$_{pph}$ has been published liganded by CMP as co-crystallized substrate.[149, 210] The lactose binding site has not been determined by bound substrate, but was well simulated based on the crystal structure analysis of an 2,3/6SiaT from *Pasteurella multocida* (2,3/6SiaT$_{pmu}$).[148] The crystal structure of 2,6SiaT$_{psp}$ has been determined both in the presence of CMP and lactose as substrates. With respect to the high homology between 2,6SiaT$_{ple}$ and 2,6SiaT$_{psp}$, the tertiary structure of 2,6SiaT$_{ple}$ was simulated using the web tool "SWISS-Model" based on 2,6SiaT$_{psp}$ as a template.[113, 114] This computer simulated structure was utilized for the active site comparison of 2,3SiaT$_{pph}$ and 2,6SiaT$_{ple}$.

```
                1                                                                75
SiaTpph     (1) ------------------------------------------------------------------MF
SiaTple     (1) MKRIFCLVSAILLSACNDNQNTVDVVVSTVNDNVIENNTYQVKPIDTPTTFDSYSWIQTCGTPILKDDEKYSLSF

               76                                                               150
SiaTpph     (3) VFCKKIFFLIFISLMILGGCNSDSKHNNSDGNITKNKTIEVYVDRATLPTIQQMTQIIN---ENSNNKKLISWSR
SiaTple    (76) DFVAPELDQDEKFCFEFTGDVDGKRYVTQTNLTVVAPTLEVYVDHASLPSLQQLMKIIQQKNEYSQNERFISWGR

              151                                                               225
SiaTpph    (75) YPINDETLLESINGSFFKNR-----PELIKSLDSMILTNEIKKVIINGNTLWAVD-VVNIIKSIEALGKKTEIEL
SiaTple   (151) IRLTEDN-AFKLNAHIYPLAGNNTSQELVDAVIDYADSKNRLNLELNTNTGHSFRNIAPILRATSSKNNILISNI

              226                                                               300
SiaTpph   (144) NFYDDGSAEYVRLYDFSRLPESEQEYKISLSKDNIQSSINGTQP-FDNSIENIYGFSQLYPTTYHMLRADIFETN
SiaTple   (225) NLYDDGSAEYVSLYNWKDTDNKSQ--KLSDSFLVLKDYLNGISSEKPNGIYSIYNWHQLYHSSYYFLRKDYLTVE

              301                                                               375
SiaTpph   (218) LPLTSLKRVLSNNIKQMKWDYFTTFNSQQKNKFYNFTGFNPEKIKEQYKASPHENFIFIGTNSGT------ATAE
SiaTple   (298) TKLHDLREYLGGSLKQMSWDTFSQLSKGDKELFLNIVGFDQEKLQQEYQQSELPNFVFTGTTTWAGGETKEYYAQ

              376                                                               450
SiaTpph   (287) QQIDILTEAKKPDSPIITNSIQGLDLFFKGHP-SATYNQQIIDAHN-MIEIYNKIPFFALIMTDALPDAVGGMGS
SiaTple   (373) QQVNVVNNAINETSPYYLG--REHDLFFKGHPRGGIINDIILGSFNNMIDIPAKVSFFVLMMTGMLPDTVGGIAS

              451                                                 505
SiaTpph   (360) SVFFSLPNTVENKFIFYKSDTDIENN-----ALIQVMIELNIVNRNDVKLISDLQ
SiaTple   (446) SLYFSIPAEKVSFIVFTSSDTITDREDALKSPLVQVMMTLGIVKEKDVLFWC---
```

Sequence 9.1. Protein sequence alignment of 2,3SiaT$_{pph}$ and 2,6SiaT$_{ple}$.

Being two functionally different sialyltransferases, understandably the protein sequences of 2,3SiaT$_{pph}$ and 2,6SiaT$_{ple}$ show only low 26.1% identity and 38.4% consensus. (Sequence 9.1) However, according to the three-dimensional structure alignment, the overall folding patterns of these two proteins present high similarity. (Figure 9.1a) Both of them have shell-shaped

tertiary structures (glycosyltransferase-B, GT-B) stabilized by a core of multi-β-pleated sheets. The active sites are located at the bottom of the shell, close to the centers of the proteins. Their active pockets are made up from the same protein domains responsible for CMP and sialyl acceptor binding, respectively. CMP is captured by a trap in closed by Thr49, Gly227, Lys315, His317, Pro318, Phe341, Glu342, Ser360 and Val361 in 2,3SiaT$_{pph}$,[148] and Ser108, Leu109, Thr342, Lys385, Gly386, His 387, Pro388, Ser412, Phe413, Glu414, Ser431, and Ser432 in 2,6SiaT$_{ple}$. Meanwhile, acceptor binding domains are Arg74, Trp121, Asp148 and Ser150 in 2,3SiaT$_{pph}$,[148] and His106, Asn157, His187 and Asp215 in 2,6SiaT$_{ple}$(Table 9.1).

Table 9.1. Corresponding residues of 2,3SiaT$_{pph}$ and 2,6SiaT$_{ple}$ for CMP and acceptor binding

| | 2,3SiaT$_{pph}$ | 2,6SiaT$_{ple}$ |
|---|---|---|
| CMP | Thr49, Gly227, Lys315, His317, Pro318, Phe341, Glu342, Ser360, Val361 | Ser108, Leu109, Thr342, Lys385, Gly386, His 387, Pro388, Ser412, Phe413, Glu414, Ser431, Ser432 |
| acceptor | Arg74, Trp121, Asp148, Ser150 | His106, Asn157, His187, Asp215 |

In both 2,3- and 2,6-SiaT, His and Asp are the key residues for the catalytic performance, in which the former acts as a catalytic acid that protonates the phosphate oxygen of CMP-sialic acid, while the latter acts as a catalytic base that deprotonates the 6-hydroxyl of the terminal Gal moiety.[148, 149] Except for this similarity, the acceptor binding domains are quite different and cause the galactopyranose moiety's retroflexion when the acceptor is entering into the active channel, exposing either the OH3 or OH6 of the galactopyranose moiety to the catalytic center in order to create and producing the products with two different regioisomeric constitutions (2,3- or 2,6-). Focusing on the entrance channel for the acceptor, 2,6SiaT$_{ple}$ exposes Trp347 on the loop1 in Figure 9.1b. Its side chain offers a huge π-electron surface forming a wall of the channel across the catalytic His106, Asn157, His187 and Asp215. (Figure 9.2a) Therefore, into this narrow channel β-galactosides can enter quite freely owing to their rather flat β-glycosidic face (Figure 9.2b). However, α-galactosides may be blocked by Trp347 during their approach to the active center, because the α-glycosidic bond is perpendicular to the galactopyranose ring, which results in an increased steric bulkiness of the acceptor molecule. This structural feature is able to explain the low activity of 2,6SiaT$_{ple}$ towards α-galactosides. On the contrary, loop 1 in 2,3SiaT$_{pph}$ is relative small. The lack of a corresponding residue for Trp347 in 2,6SiaT$_{ple}$ offers more space for approaching acceptors. Both α- and β-galactosides fit quite well in positions close to the CMP location. Thus, only little activity difference of 2,3SiaT$_{ple}$ was determined when methyl-α- and β-D-galactopyranosides were tested as acceptors.

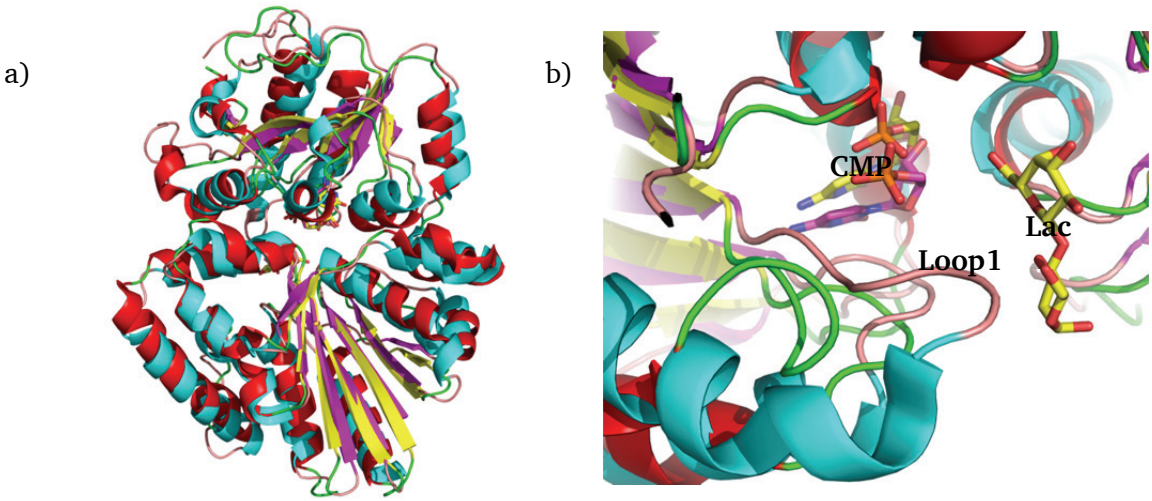



Figure 9.1.Three-dimensional structure alignment of 2,3SiaT$_{pph}$ and 2,6SiaT$_{ple}$. a) Overall structures alignment of 2,3SiaT$_{pph}$ (PDB ID 2ZWI) and 2,6SiaT$_{ple}$. b) Active domain alignment of 2,3SiaT$_{pph}$ and 2,6SiaT$_{ple}$. Red-green-yellow stands for 2,3SiaT$_{pph}$, flesh red-acid blue-pink stands for 2,6SiaT$_{ple}$.

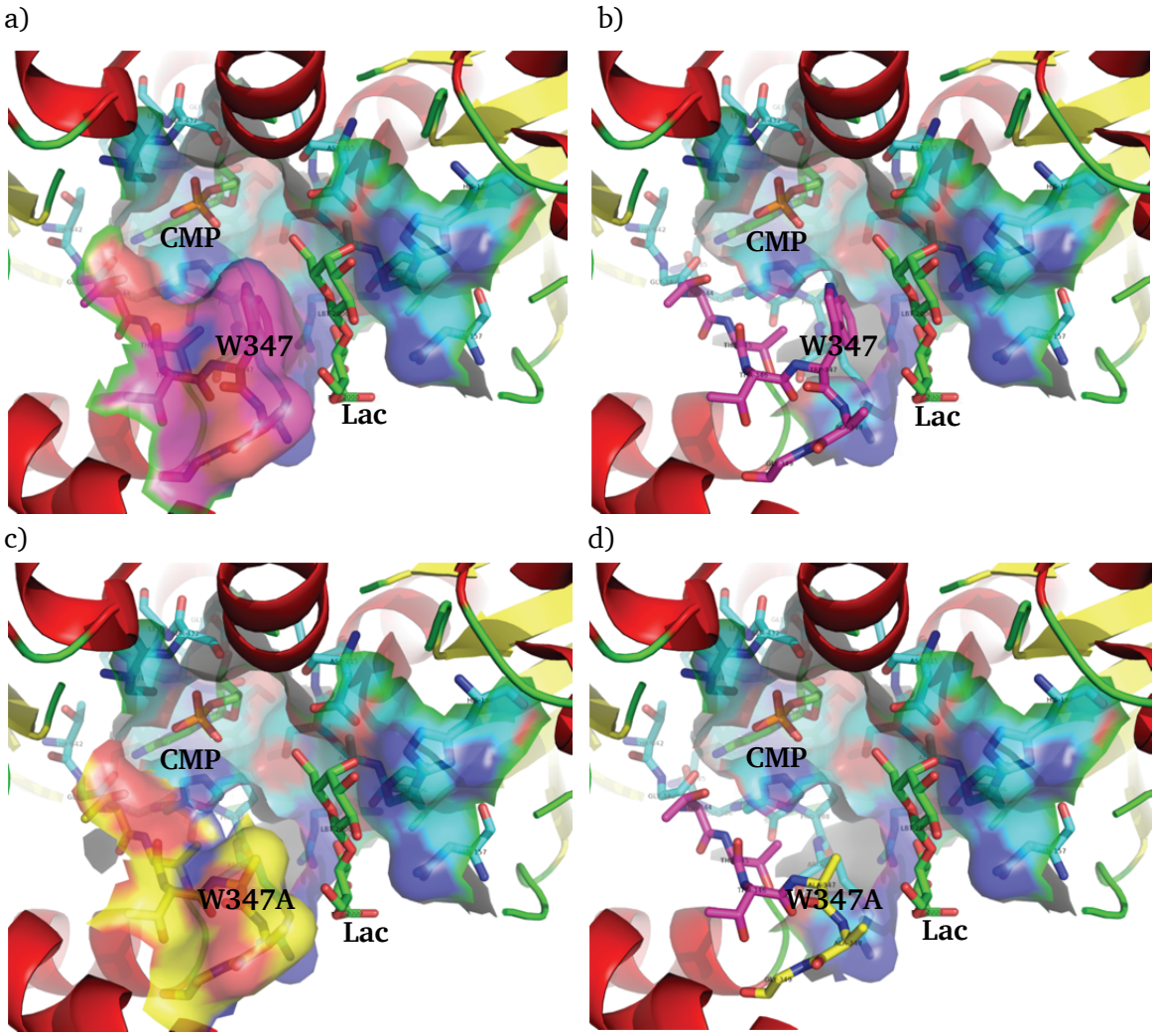



Figure 9.2. Structure analysis of active domain. a) Surface of loop1 in 2,6SiaT$_{ple}$. b) Amino acid residues of loop1 (pink) in 2,6SiaT$_{ple}$. c) Surface of loop1 in W347A of 2,6SiaT$_{ple}$. d) Amino acid residues of loop1 in W347A of 2,6SiaT$_{ple}$.

In order to improve the tolerance of 2,6SiaT$_{ple}$ for α-galactoside, the modification has to consider the most discriminating Trp347 as well as its spatical neighbors located on the same loop, including Thr344, Thr345, Thr346, Trp347, Ala348 and Gly349. Building a random mutagenesis library covering the whole loop would be an ideal way to test all amino acid combinations for activity. However, too many variants would have to be picked, which makes the screening a formidable task. Thus, the Trp347 site was firstly focused on to see whether it can be exchanged for other residues (named Lib-W347). Furthermore, replacing Trp347 into non-polar hydrophobic amino acid residues such as Ala or Gly could potentially eliminate the block for an entrance of α-galactosides (Figure 9.2a and b). Unfortunately, the W347A variant has been reported to loosing SiaT activity almost completely.[149] Therefore, if Trp347 needs to be changed into Ala or Gly, a similar residue would be required to activity recover somewhere else on the loop. Upon inspection of this loop, Thr344, Thr345 and Thr346 are oriented towards CMP and the galactopyranose ring of the acceptor, which makes them potential candidate sites for activity recovering mutagenesis. However, the side chain orientations of Thr344 and Thr346 point away from the reaction center between the phosphate group of CMP and the galactopyranose moiety of the acceptor, and only the side chain of Thr345 is directed to CMP. Therefore, Thr345 was chosen on rational consideration as the most promising mutagenesis site. On the other hand, Gly can sustain a larger turn angle then Ala in the tertiary structure of the protein. According to the analysis above, both Trp347Ala and Trp347Gly site mutation were tried to be induced into 2,6SiaT$_{ple}$ first, followed by a building of saturation mutagenesis library at Thr345 (named Lib-W347A-T345 and Lib-W347G-T345). In summary, it was planned to construct three mutagenesis libraries with different randomization sites, which are Lib-W347, Lib-W347A-T345 and Lib-W347G-T345.

### 9.2.2. Building and screening of 2,6SiaT$_{ple}$ mutagenesis libraries

For Lib-W347A-T345 and Lib-W347G-T345, single site mutations W347A (GGG) and W347G (GCG) were firstly introduced into 2,6SiaT$_{ple}$. The gene identification shows that the wild-type codon TGG (Trp) has been successfully replaced by GGG (Ala) and GCG (Gly), respectively (Figure 9.3a). For Lib-W347, primers with NNS codon degeneracy at W347 site were designed for PCR. For Lib-W347A-T345 and Lib-W347G-T345, similar NNS codon degeneracy primers

were also designed based on the mutate sequence as templates. The mutation frequency of all three libraries was evaluated by sequencing of mixed plasmids presented in Figure 9.3.

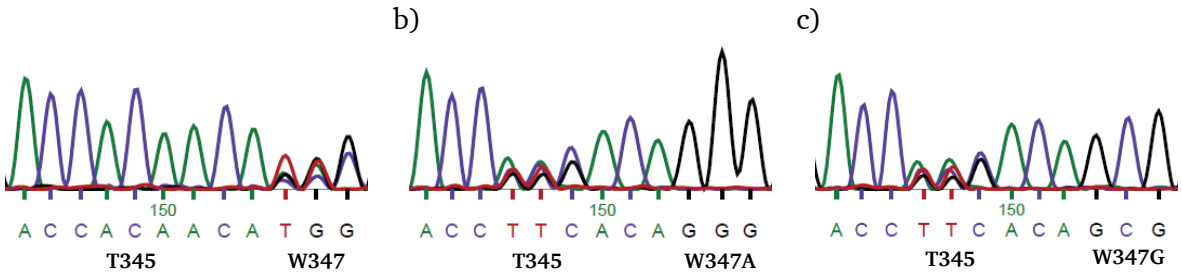



Figure 9.3.Evaluation of 2,6SiaT$_{ple}$ mutagenesis libraries. a) Lib-W347. b) Lib-W347A-T345, c) Lib-W347G-T345. The original codon on W347 or T345 has been successfully replaced by NNS, in which N was shown as A/T/G/C overlapping peaks and S as G/C overlapping peaks.

In Chapters 4 and 8, pH-Based high throughput assay methods have been successfully applied already for screening of TK and CSS mutant libraries where serveral promising mutants with extended substrate tolerance could be identified. Similar to those, as well as to studies involving the modification of acceptors for sialyltransferase, the release of protons during the reaction can also be determined spectrophotometrically by assistance of a pH indicator. According to this assay principle, the substrate tolerance and kinetic constants of 2,3SiaT$_{pph}$ and 2,6SiaT$_{ple}$ where already determined in Chapter 7. Because of its high sensitivity and high-throughput nature, this assay can also be utilized for the screening of SiaT libraries. The screening reaction system was basically the same as that for determination of SiaT substrate tolerance. Phenol red at (2.82 mM concentratrion) was used as pH indicator, and 2 mM Tris-HCl buffer (8.0) with 0.5 M NaCl was chosen as buffer system. Substrates 0.5 mM CMP-Neu5Ac and 50 mM methyl-α-D-galactopyranoside were added as donor and acceptor, respectively. The cultivation and lysis methods for cells expressing the SiaT variant followed the same protocals as described in Chapter 4/8 for TK/CSS mutagenesis libraries. As the enzyme sample, 40 $\mu$l each of crude extract was used.

All of the 6 plates in the three libraries were screened for twice. Unfortunately, Lib-W347A-T345 and Lib-W347G-T345 gave no positive mutants at all. That means the activity in the 2,6SiaT$_{ple}$-W347A and –W347G variants cannot be recovered by any site-mutation at the T345 position. Interestingly, four colonies in Lib-W347 presented positive assay results. However, sequencing upon all hits were confirmed still to be wild-type. This disappointing screening result reveals that W347 plays an extremely crucial role in the 2,6-sialyltransfer reaction, which cannot be replaced by any other amino acid residues at the same position or be rebuilt at the corresponding alternative Thr345 site.

### 9.2.3. Re-analysis of active site of 2,6SiaT$_{ple}$

Although there was no anticipated positive result coming out from the library screening, it still provided certain information about the catalytic mechanism of 2,6-SiaT and the function of Trp347 in the catalytic center. Kakuta *et al.* have interpreted the proposed catalytic mechanism of 2,6SiaT.[149] It is considered that in 2,6SiaT$_{ple}$ Asp215 deprotonates the OH6 of the galactopyranose moiety of the acceptor, thereby increasing its nucleophilic character for the attack on C2 of the NeuAc moiety. On the other hand, His387 protonates the phosphate oxygen of CMP-NeuAc stabilizing the oxocarbenium-ion-like transition state, followed by the release of CMP and the formation of the product.[149] Trp347 seems not to be involved in the generation of the transition state. On the other hand, 2,3SiaT shares the same catalytic mechanism with 2,6SiaT,[148] but has no residue corresponding to Trp. This might indicate that the Trp moiety is not required for the sialyltransfer reaction but, irrespective of the results from site-directed mutagenesis[148, 150] or from the randomized mutagenesis libraries constructed here, Trp shows a unique function in 2,6SiaT. Ni *et al.* have proved that Trp270 in 2,3/6SiaT$_{pmu}$ interacts with the galactopyranose ring by hydrophobic stacking.[150] The $K_M$ change of W270F and W270A further showed that Trp270 binds to both donor and acceptor.[150] Therefore, it is concluded that Trp347 in 2,6SiaT$_{ple}$ also very likely not only engages in hydrophobic stacking interaction with the acceptor galactopyranose ring, but also contributes to the binding both of donor and acceptor by acting as a bridge between them. This bridge seems to stabilize both the Neu5Ac moiety release from CMP-Neu5Ac and the attack of O6 of galactopyranose moiety to C2 of the Neu5Ac unit. Moreover, the retroflexion of the galactopyranose moiety causes the distinct regiospecific product formation characteristic for by 2,3- and 2,6-SiaTs. Thus, we also think that the presence of Trp347 in the 2,6SiaT$_{ple}$ helps to adjust the correct conformation of lactose to favor the formation of a 2,6-sialated product.

## 9.3. Conclusion

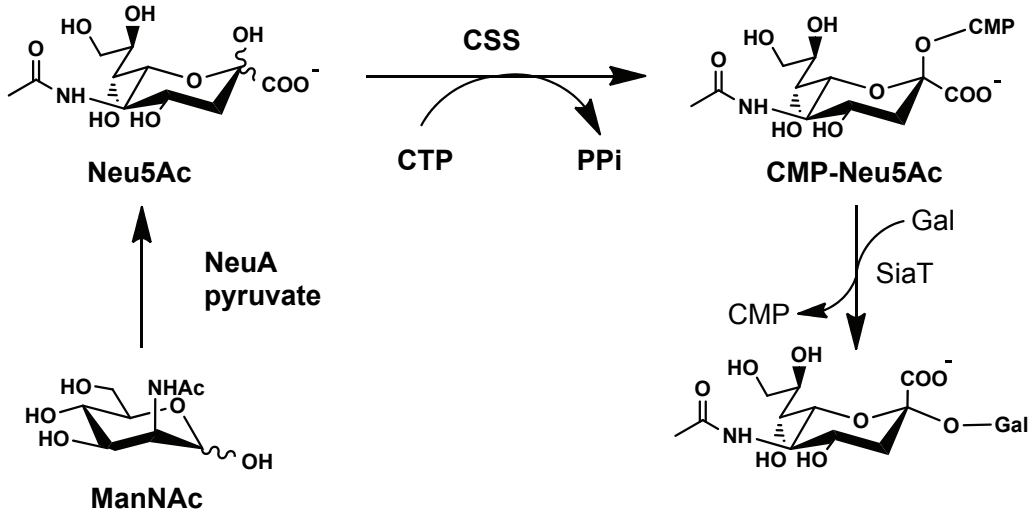
Different from 2,3SiaT$_{pph}$, 2,6SiaT$_{ple}$ can only accept β-galactosides well. From the structure alignment between these two SiaTs, Trp347 was found as the crucial residue which blocks the entrance for α-galactosides to the active center. Therefore, three mutagenesis libraries were

constructed to improve the activity of 2,6SiaT$_{ple}$ towards α-galactosides to replace or relocate Trp347. An optimized pH-based assay was developed as a high-throughput screening method using methyl-α-D-galactopyranoside as an acceptor. However, no positive mutant could be detected during the screening, which indicates that Trp347 seems to be irreplaceable in 2,6SiaT$_{ple}$. It is presumed that Trp347 contributes to the binding between donor and acceptor and helps to stabilize the correct acceptor conformation to form the 2,6-sialated product.

## 10. Chemoenzymatic Synthesis of neo-2,3- and 2,6-Sialoconjugates

### 10.1. Introduction

The synthesis of neo-sialoconjugates *in vitro* is a chemoenzymatic pathway, which contains four steps: chemical synthesis of Man(NAc) derivatives, followed by enzymatic aldolization, CMP-activation and sialylation. The aldolase catalyzed carboligation produces sialic acids by starting from Man(NAc) or analogs and pyruvate. The CMP-activation of sialic acids is catalyzed by CMP-sialate synthetase forming the donor required for the finial sialyltransfer step which is catalyzed by a 2,3SiaT or 2,6SiaT to produce the target 2,3- or 2,6-sialoconjugates (Scheme 10.1).[130, 171] Ideally, the three enzymatic steps can be carried out together in a one-pot approach.[202]




Scheme 10.1. The enzyme-catalyzed synthetic route to sialcoconjugates

The diversity of the synthesized sialoconjugates is determined by the regioselectivity of the chosen SiaT enzyme and by the potential diversity of the employed starting donor and acceptor in the sialyltrasfer. Although the natural acceptors of sialylation are mostly limited to a Gal or GalNAc moiety, the structure of artificial galactosides can be diverse, for example, the

C2 modified Gal or GalNAc, C1 modified lactose and LacNAc, and even modified, labeled or complex immobilized oligosaccharides such as Lewis$^X$ in the potential research on medical purposes.[127, 147, 202, 204, 219-223] In addition, the extremely broad acceptor specificity of 2,3SiaT from *Photobacterium* sp. even allows sialyltransfer to related acceptors like α-Glc, β-Man and inositol, thereby producing unique sialyloligosaccharides.[224, 225]

Up to present, besides natural sialic acids, many non-natural analogs with modification at C5, C7, C8 and C9 have been linked to a fluorescent Lac with α-2,6-sialyl linkage.[171] In this chapter, six typical sialic acids and their derivatives were transferred to the same fluorescent Lac acceptor with an α-2,3-linkage by 2,3SiaT$_{pph}$, and a fluorescent LacNAc acceptor with an α-2,6-linkage by 2,6SiaT$_{ple}$, respectively.

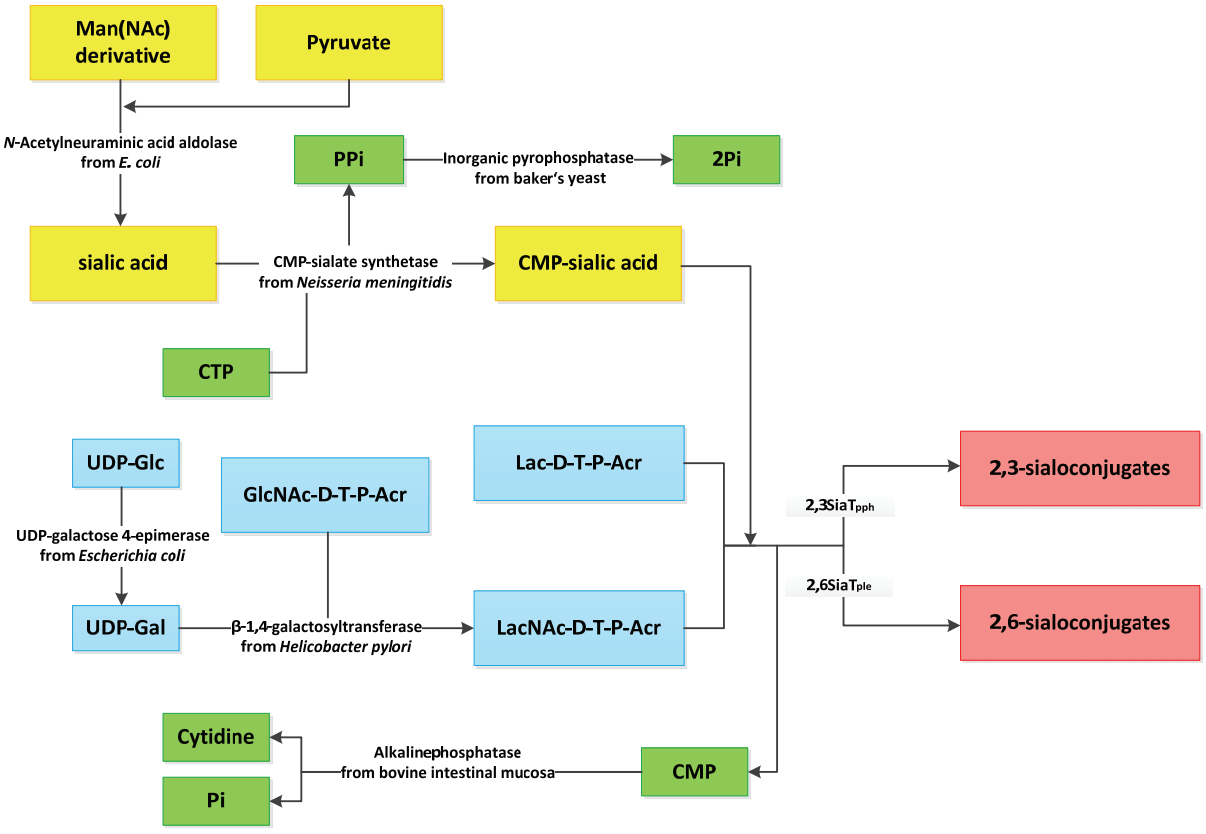
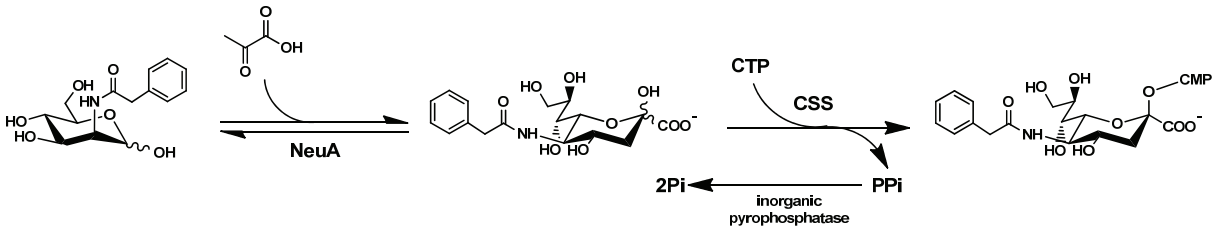



Figure 10.1. Chemoenzymatic synthetic route for the synthesis of fluorescent sialoconjugates

## 10.2. Results and discussion

### 10.2.1. Enzymatic synthesis and CMP activation of sialic acid derivatives

The largest group of structureal variations in sialic acid is mainly due to modification of C5 and C9 of Neu5Ac. For the present synthesis study, Neu5Ac, Neu5Gc, KDN, epi-KDN, Neu5Ac-9-$N_3$ and Neu5NPhAc were chosen as typical sialic acid donors. Besides Neu5Ac, Neu5Gc and KDN are also the most common sialic acids in nature. Neu5Gc has an additional hydroxyl group at the acetamide group of Neu5Ac. Accordingly, Neu5NPhAc contains a phenyl group at the same position, which is non-polar, much larger than the hydroxyl group and which can be specifically removed by Penicillin G acylase (PGA). Kdn has a hydroxyl group instead of the acetamide group of Neu5Ac. Oppositely, *epi*-Kdn contains the same hydroxyl in the epimeric *trans* configuration. In addition, Neu5Ac-9-$N_3$ brings in a modification on C9 of Neu5Ac. Based on these six sialic acids, the donor specificity of SiaT can be tested for various structural variations along the chain of sialic acid, as well as the configuration of C5. The synthesis of these sialic acids starts from the corresponding Man(NAc) derivatives and pyruvate, and are catalyzed by *N*-acetylneuraminic acid aldolase (NeuA) from *E. coli* (Scheme 10.2). This aldolization reaction is a reversible reaction, which leads to a product yield around 50-60%. Therefore, excessive pyruvate was added to the reaction mixture in order to increase the product fraction at equilibrium.




Scheme 10.2. Examplary synthesis of Neu5NPhAc and its CMP-activation.

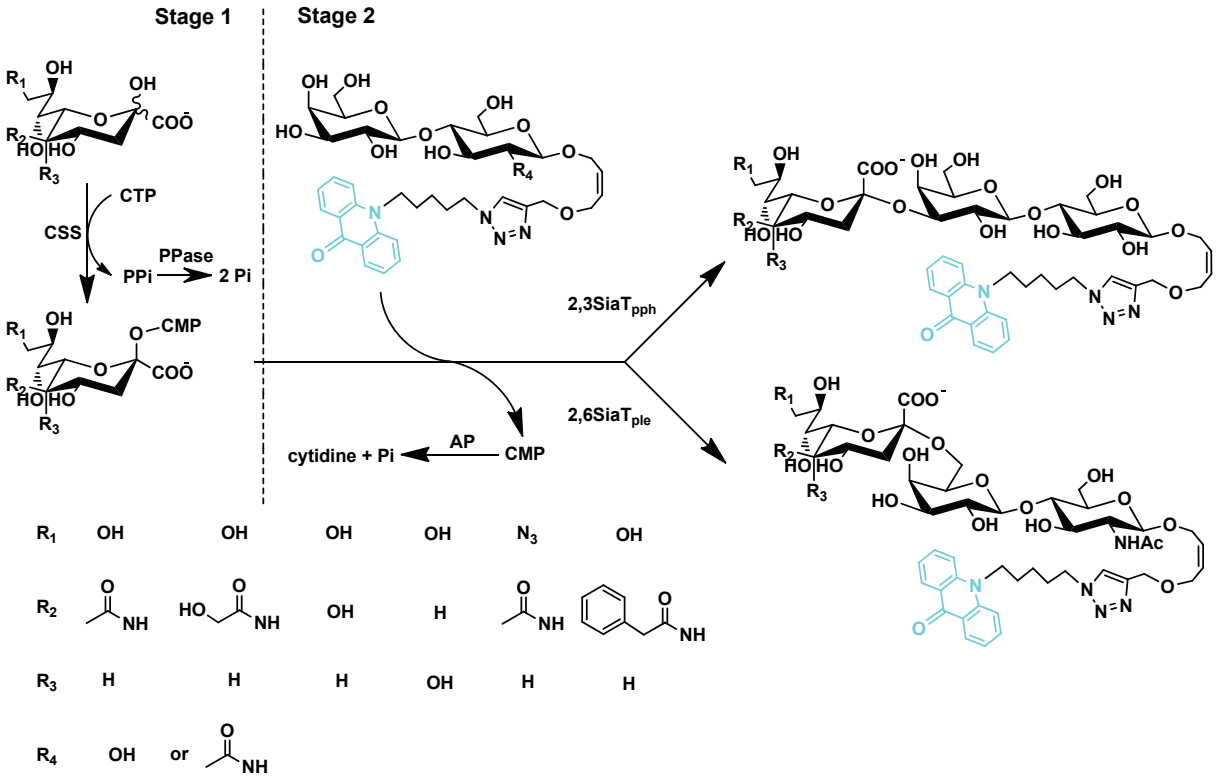
The CMP activation of sialic acid can be carried out easily together with the sialylation in a one pot reaction system.[171] Alternatively, it can also be performed separately as a single step in the synthesis process. For the determination of kinetic constants of SiaT, CMP-Neu5Ac was synthesized and isolated as a purified compound. The activation of Neu5Ac could be

completed with 100% conversion. Because of the relative instability of CMP-Neu5Ac in acidic solution, the purification of CMP-Neu5Ac on a Bio-gel column was performed with dilute aqueous ammonia (pH 9.0) as mobile phase to reach a final yield of 90%. On the other hand, large hydrophobic the phenyl modification in Neu5NPhAc leads to a significantly lower acceptance by $CSS_{nme}$. In chapter 8, the mutation of $CSS_{nme}$ at Phe192 or/and Phe193 is shown to lead to variants having improved activity towards 5' modified sialic acids. Therefore, the CMP-activation of Neu5NPhAc was carried out preferentially by $CSS_{nme}$-F192A instead of the wild-type $CSS_{nme}$, which was used for the other sialic acids in a one-pot reaction by coupling with SiaT.

### 10.2.2. Enzymatic synthesis of neo-2,3-sialoconjugates

The preparative regioselective synthesis of neo-sialoconjugates by using $2,3SiaT_{pph}$ and $2,6SiaT_{ple}$ was carried out by a one-pot two-enzyme approach, in which $CSS_{nme}$ and the correspongding SiaT were involved. Additionally, inorganic pyrophosphatase from baker's yeast was added into the CMP-activation step to eliminate the product inhibition of pyrophosphate (PPi) for $CSS_{nme}$. When the CMP-activation was completed, alkaline phosphatase from bovine intestinal mucosa was involved in the sialyltransfer step to eliminate the inhibition of released CMP for SiaT. With these two auliliary enzymes, CMP-activation and sialyltransfer can be performed much faster and closer to complete. The 2,3SiaT from *Photobacterium* sp. has been demonitrated before to have an extremely broad substrate range, which was confirmed by the pH-based assay in chapter 7. Actually, 2,3SiaT from *Photobacterium* sp. is able to accept all of the positions having an equatorial hydroxyl group of a *cis*-diol structure stereoconfigurationally corresponding to the C3 and C4 unit of the Gal moiety.[224] Moreover, the fluorescant label on C7 of Lac did not decrease the activity of SiaT.[171] Here, the synthesis of neo-2,3-sialoconjugates also used the fluorlabeled Lac acceptor (Lac-D-T-P-Acr). During the sialyltransfer reaction, $2,3SiaT_{pph}$ can accept all of the six CMP-activated sialic acids. TLC monitoring of the reaction showed that Neu5Ac, Neu5Gc, Kdn and Neu5Ac-9-$N_3$ all could be quantitatively transferred into the Gal moiety of Lac-D-T-P-Acr. The final isolated product yields were 84%, 77%, 86% and 82%, respectively. However, the transfer rates for *epi*-Kdn and Neu5NPhAc were sloer and reached only around 60% and 50% conversion under non-optimized conditions. As a result, the total yields of *epi*-KDN-2,3-Lac-D-T-P-Acr and Neu5NPhAc-2,3-Lac-D-T-P-Acr were only 58% and 48%. This result reveals that the structure and configuration of the C5 side chain is more critical to the activity of CSS

and 2,3SiaT$_{pph}$ compared with C9 substitution. The modification with *trans* configuration or a large hydrophobic amide group can obviously decrease the activity and conversion efficiency of 2,3SiaT$_{pph}$.




Scheme 10.3. Enzymatic synthesis of neo-2,3- and 2,6-sialoconjugates.

### 10.2.3.Enzymatic synthesis of neo-2,6-sialoconjugates

Compared with 2,3SiaT$_{pph}$, the 2,6SiaT$_{ple}$ has a much narrower acceptor range, because the Trp347 at the entrance of its active pocket blocks the access for all α-galactosides and provides the enzyme with a stricter β-selectivity for sialylation acceptors. Fortunately, however, 2,6SiaT$_{ple}$ was found to tolerate a large donor range, which rendered the synthesis of the corresponding 2,6-sialylated Lac-D-T-P-Acr acceptor successful.[171] The structurally related LacNAc is the terminal moiety in the Lewis$^x$ structure, where it plays the role of the natural acceptor for SiaT. Thus, the acridone conjugate LacNAc was synthesized in chapter 8 to be used for the neo-2,6-sialoconjugate synthesis. Using the same one-pot enzyme cascade as above, the six sialic acids were activated, then transferred by 2,6SiaT$_{ple}$ into the Gal moiety of LacNAc-D-T-P-Acr, thereby forming the expected 2,6-LacNAc-sialoconjugates with product

yields from 51% to 86%. Different from 2,3SiaT$_{pph}$, the 2,6SiaT$_{ple}$ can accept well not only Neu5Ac, Neu5Gc, Kdn and Neu5Ac-9-$N_3$, but also *epi*-Kdn with close to 100% conversion rate (by TLC control) and 86% final yield of its product. Only Neu5NPhAc could be transferred less effectively by 2,6SiaT$_{ple}$, showing a relative conversion rate of around 50% and a total yield of 51% Neu5NPhAc-2,6-LacNAc-D-T-P-Acr. On the whole, 2,6SiaT$_{ple}$ seems to have a similarly broad donor specificity like 2,3SiaT$_{pph}$, except for its apparent higher tolerance for an inverted configuration at the C5 position like for 2,3SiaT$_{pph}$, the *N*-phenylacetyl modification in Neu5NPhAc decreases the affinity and activity of 2,6SiaT$_{ple}$.

## 10.3. Conclusion

A series of neo-2,3- and 2,6-sialoconjugates were synthesized in a one-pot, two-enzyme approach using a combination of CSS$_{nme}$ with 2,3SiaT$_{pph}$ or 2,6SiaT$_{ple}$. Six typical sialic acids (Neu5Ac, Neu5Gc, Kdn, *epi*-Kdn, Neu5Ac-9-N3 and Neu5NPhAc) with structural modifications at C5 and C9 were involved as donors for a transfer to Lac-D-T-P-Acr and LacNAc-D-T-P-Acr with α-2,3- and α-2,6-configuration, respectively. Both 2,3SiaT$_{pph}$ and 2,6SiaT$_{ple}$ showed quite similar donor specificity towards these six sialic acids. Both of them have high activity towards CMP activated Neu5Ac, Neu5Gc, Kdn and Neu5Ac-9-$N_3$, but limited tolerance for the bulky Neu5NPhAc. As a clear distinction, the 2,6SiaT$_{ple}$ can accept *epi*-Kdn better than 2,3SiaT$_{pph}$.

# 11. Cloning, Expression and Characterization of a Potential 2,3-Sialyltransferase from *Shewanella piezotolerans*

## 11.1. Introduction

Finding novel and superior biocatalysts is a top priority in biocatalysis. While more and more genome sequence data is being added into gene databases, sequence alignment has become an effective tool to search for potential functional proteins from Genbank. Meanwhile, gene synthesis technique has been highly developed, so that gene sequences coding for desired enzymes can be artificially synthesized at relatively low cost. This technique can be applied widely in cloning, expression and further characterization of potential proteins without the need for growing of the resource organism, which is especially important for difficult to culture microorganisms. Recently, a new potential sialyltransferase from *Shewanella piezotolerans* (SiaT0160, $SiaT_{spi}$) has been identified from genome sequencing[201] (Genbank Access No. CP000472) and assigned to the GT80 family (www.cazy.org) because of its high homology with α-2,3SiaTs in the same family. However, cloning, expression and activity assay for this $SiaT_{spi}$ have not been reported yet. In this chapter, an artificial gene of this protein is cloned, modified and expressed in *E. coli* and a *Yarrowia* expression system. The primary 2,3SiaT function of this enzyme is characterized.

## 11.2. Results and discussion

### 11.2.1. Sequence and structure analysis of $SiaT_{spi}$

#### 11.2.1.1. Protein sequence analysis of $SiaT_{spi}$

*S. piezotolerans* is a marine microorganism living in the deepsea.[226] The genomic DNA sequence of *S. piezotolerans* has been sequenced[201] and indexed in Genbank (Access No. CP000472). Among its whole genome sequence, an open reading frame has been identified

for which the corresponding protein expression product has high homology with α/β-2,3 and α-2,6SiaTs belonging to the GT 80 family (www.cazy.org) (Sequence 11.1). From the protein sequence alignment, it is shown that the protein is more similar to 2,3SiaT than to 2,6SiaT. For example, it has 27.9% identity and 42.9% consensus with 2,3SiaT$_{pph}$, but only 19.3% identity and 31.4% consensus with 2,6SiaT$_{ple}$. From the sequence feature, it is equipped with both the typical CMP and lactose binding domains, which are also very similar to those of 2,3SiaT$_{pph}$. Thus, it is assumed that SiaT$_{spi}$ possibly has a 2,3SiaT function.


Sequence 11.1. Protein sequence alignment of SiaT$_{spi}$ and SiaT$_{spi}$-U with 2,3SiaT$_{pph}$ and 2,6SiaT$_{ple}$. Red frames indicate upstream sequence and 12 amino acid insertion in the middle of SiaT$_{spi}$.

However, an interesting observation is that the gene of SiaT$_{spi}$ listed in the Genbank and the CAZy database contains only 1,101 bp encoding a predicted protein of 367 amino acid residues, which is significantly shorter than other SiaTs in the same family (usually around 400 amino acid residues for 2,3SiaT and 500 amino acid residues for 2,6SiaT), while this

might be the result of evolution to adapt the organism to its preferred cold and pressurized habitat, it could also be due to a mistake generated by the automated sequence analysis software which might have misinterpreted the start codon ATG of the open reading frame. Using BLAST, indeed a missed sequence fragment of 96 bp can be identified. This sequence fragment codes for an additional series of amino acid residues that shows high similarity with the corresponding part of other SiaTs (Sequence 11.1, red frame 1). When this sequence is joined to the $SiaT_{spi}$, $SiaT_{spi}$ reaches the full sequence length typical for other 2,3SiaTs. Thus, it is inferred that this sequence might be an essential functional part of $SiaT_{spi}$. On the other hand, a corresponding 24 amino acid sequence at the N-terminus of $2,3SiaT_{pph}$ has been identified as a common signal peptide sequence, which does not contribute to the sialyltransfer function of the enzyme.[161]

As an organism living in the deepsea, *S. piezotolerans* has to adapt its metabolism and constituents to the cold and pressurized environment. Therefore, the $SiaT_{spi}$ might be likely a cold-adapted enzyme. Normally, cold-adapted enzymes have remarkable sequence signatures. For example, they contain lower numbers of polar (Ser, Thr, Asn, Glu and so on) and aromatic (Tyr, Phe and Trp) amino acid residues to reduce the number of stabilizing hydrogen bonds and aromatic π- and hydrophobic interactions.[227, 228] The number of Pro residues is also reduced because of the rigidity of its imino group.[227] Gly stacking may be found near the active site because its absent side chain offers varied dihedral angle in protein conformation, which is beneficial for the flexibility of the enzyme.[227] Because the guanidinium group of Arg which can form at least one salt bridge in the tertiary structure, a lower number of Arg residues or a reduced ratio of Arg/(Arg+Lys) is often can be observed in thermolabile enzymes.[227]

To find out whether $SiaT_{spi}$ fits the features of a cold-adapted enzyme, its amino acid residue composition and secondary structure were analyzed using the $2,3SiaT_{pph}$ from a mesophilic source as reference. The analysis data is listed in Table 11.1. Compared to $2,3SiaT_{pph}$, the percentage of polar amino acid residues is obviously reduced. On the other hand, the total number of aromatic residues is almost the same as in $2,3SiaT_{pph}$, and both of them contain the same number of Pro residues. The Gly content in $SiaT_{spi}$ is slightly higher than that found in $2,3SiaT_{pph}$. All these data show that $SiaT_{spi}$ may have a lower thermostability than $2,3SiaT_{pph}$. However, the doubled number of Arg and ratio Arg/(Lys+Arg) suggests the opposite conclusion. Considering the relative low optimum temperature of $2,3SiaT_{pph}$ (30°C),[161] the analysis for which partly fits the property of a cold-adapted enzyme itself,[229] the $SiaT_{spi}$ could thus also be considered as a cold-adapted enzyme.

Table 11.1. Amino acid residue composition and secondary structure analysis of $SiaT_{spi}$ and $2,3SiaT_{pph}$

| | $SiaT_{spi}$ | $SiaT_{spi}$-U | $dSiaT_{spi}$-U | $2,3SiaT_{pph}$ |
|---|---|---|---|---|
| Total residues | 367 | 399 | 387 | 409 |
| Ser | 9.68% | 9.52% | 9.82% | 8.31% |
| Thr | 3.72% | 3.76% | 3.26% | 6.36% |
| Asn | 5.71% | 5.76% | 5.43% | 10.02% |
| Glu | 4.42% | 4.26% | 4.39% | 6.11% |
| Tyr | 4.47% | 4.51% | 4.65% | 3.64% |
| Phe | 3.97% | 4.01% | 4.13% | 6.36% |
| Trp | 1.24% | 1.25% | 1.29% | 0.73% |
| Pro | 3.54% | 3.76% | 3.88% | 3.67% |
| Gly | 5.18% | 4.76% | 4.65% | 4.16% |
| Arg | 3.72% | 3.76% | 3.62% | 1.96% |
| Arg/(Lys+Arg) | 41.7% | 40.5% | 40.0% | 20.5% |
| Disulfide linkage | none | none | none | none |

#### 11.2.1.2. Analysis of the three-dimensional structure of $SiaT_{spi}$

The crystal structure of $2,3SiaT_{pph}$ has been determined containing CMP as co-crystallized substrate part.[148] Using this crystal structure as a template, the tertiary structure of $SiaT_{spi}$ with and without the identified additional upstream sequence has been simulated using the web-tool SWISS Model,[113, 114] to evaluate its active site and analyze the potential function of its upstream sequence. SWISS Model uses protein sequence alignment and known crystal structure databases (e.g. PDB) to calculate a model of the unknown protein structure.[113, 114] Because of the high homology between $SiaT_{spi}$ or $SiaT_{spi}$-U and $SiaT_{pph}$, the latter was chosen as the template for automatic structure simulation of $SiaT_{spi}$ by SWISS Model. Figure 11.1 shows a comparison of the structures of $2,3SiaT_{pph}$, $SiaT_{spi}$ and $SiaT_{spi}$-U. A high similarity can found from the structure overlay. In the $2,3SiaT_{pph}$ structure, 7 parallel β-

sheets form a stable support motif. In comparison, however, in $SiaT_{spi}$ one large β-sheet and one α-helix are missing at the N-terminus when its upstream sequence is missing. This void may lend more flexibility to $SiaT_{spi}$ when needed to adapt to a cold environment under high pressure. If the upstream sequence is added to $SiaT_{spi}$, these missing structures are fully reconstructed. Because the active site of 2,3SiaT is located in the center of the protein structure, this upstream sequence part could play an important role in the activity and stability of this potential 2,3SiaT. Therefore, it is believed the upstream sequence contributes to the activity of $SiaT_{spi}$, at least within a normal environment. On the other hand, when compared to other SiaTs, $SiaT_{spi}$ contains an additional 12 amino acid residue insert in the middle of its protein sequence. These 12 residues are interpreted to create a double β-sheet structure on one site of $SiaT_{spi}$, which may give more stability for $SiaT_{spi}$. Although there is no obvious change to the overall tertiary folding pattern induced by these two β-sheets, this might affect the activity of $SiaT_{spi}$. To investigate these conjectures, the 3 gene variants of (1) $SiaT_{spi}$, (2) $SiaT_{spi}$ with upstream sequence ($SiaT_{spi}$-U) and (3) $SiaT_{spi}$-U with deletion of the 12 residue sequence ($dSiaT_{spi}$-U) were constructed and cloned into *E. coli* BL21 cells using pET21a and pET19b as expression vectors. The expression of these three recombinant enzymes (with His-Tag at either N or C terminus) is described below (Figure 11.1).

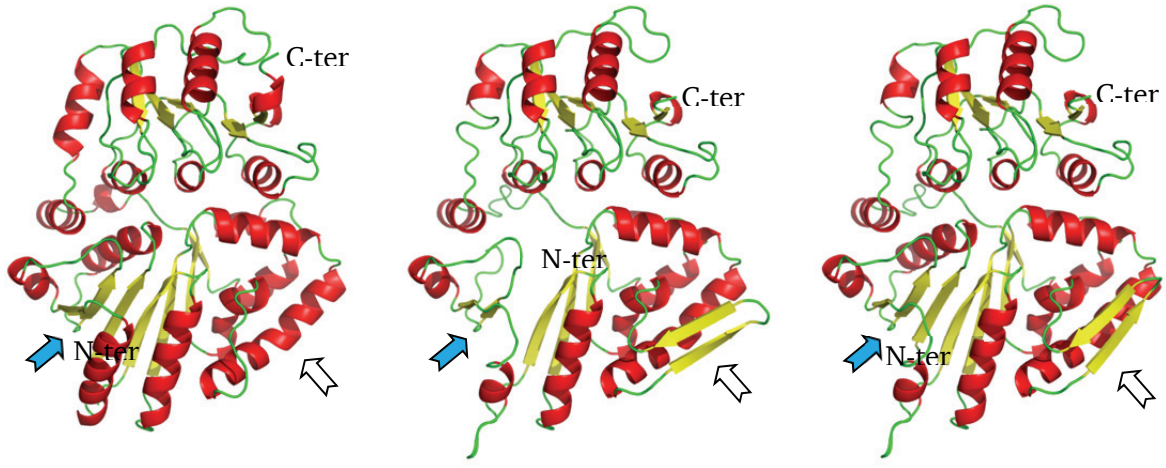



Figure 11.1. Amino acid sequence alignment and structure simulation of $2,3SiaT_{pph}$ (PDB code 2ZWI), $SiaT_{spi}$ and $SiaT_{spi}$-U by SWISS Model.[113, 114] Blue arrow: the missing upstream sequence forms two β-sheets. White arrow: the 12 amino acid insection forms two β-sheets.

The structure simulation of the active pocket could possibly directly reveal the enzyme's catalytic property. The active site of $SiaT_{spi}$ was aligned with those of $2,3SiaT_{pph}$ and $2,3/6SiaT_{pmu}$ for an analysis of the CMP and lactose binding residues, respectively (Figure

11.2). The CMP binding domain of $SiaT_{spi}$ is exactly matching that of $2,3SiaT_{pph}$. It is composed of Thr23, Gly265, Lys301, His303, Pro304, Ile325, Pro 326, Phe327, Glu328, Ser345 and Ser346, corresponding to Thr49, Gly277, Lys315, His317, Pro318, Ile339, Pro340, Phe341, Glu342, Ser359, and Ser360 in $2,3SiaT_{pph}$.[148] However, the lactose binding site is somewhat different between $SiaT_{spi}$ and $2,3SiaT_{pph}$. Both of them have the key catalytic Asp residue (Asp122 in $SiaT_{spi}$ and Asp148 in $2,3SiaT_{pph}$) and a galactopyranose moiety binding Trp residue (Trp95 in $SiaT_{spi}$ and Trp121 in $2,3SiaT_{pph}$), but $SiaT_{spi}$ contains Tyr94 instead of Leu120 in $2,3SiaT_{pph}$ and an additional Arg176, which may serve for hydrogen bonding to a glucopyranose moiety. Moreover, both $SiaT_{spi}$ and $2,3SiaT_{pph}$ share the same loop 1 on the CMP site without a Trp constituent, which may act as an irreplaceable role for 2,6SiaT activity. This could be taken as further evidence that $SiaT_{spi}$ is a potential 2,3SiaT.

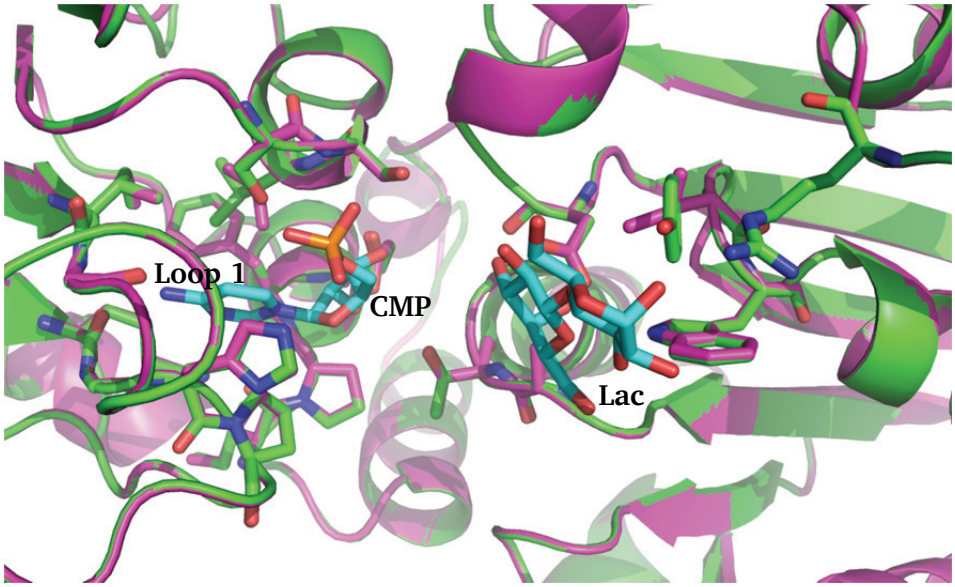



Figure 11.2. Active site alignment of $SiaT_{spi}$ (green) and $2,3SiaT_{pph}$ (pink, PDB code 2ZWI).

Table 11.2. Corresponding residues of $SiaT_{spi}$ and $2,3SiaT_{pph}$ for CMP and acceptor binding

| | $SiaT_{spi}$ | $2,3SiaT_{pph}$ |
|---|---|---|
| CMP | Thr23, Gly265, Lys301, His303, Pro304, Ile325, Pro 326, Phe327, Glu328, Ser345, Ser346 | Thr49, Gly277, Lys315, His317, Pro318, Ile339, Pro340, Phe341, Glu342, Ser359, Ser360 |
| acceptor | Asp122, Trp95 | Asp148, Trp121 |

## 11.2.2. Cloning, expression and characterization of $SiaT_{spi}$, $SiaT_{spi}$-U and $dSiaT_{spi}$-U in *E. coli* expression system

The natural habital of *S. piezotolerans*[226] is quite hard to be simulated in the lab. Therefore, artificial gene synthesis was chosen to obtain a more convenient access to the $SiaT_{spi}$ gene. The DNA sequence of the $SiaT_{spi}$-U gene was optimized for expression in *E. coli* and synthesized commercially. Restriction sites *Nde*I and *Bam*HI were added at the 5' and 3' terminus, respectively, to facilitate further cloning work. The shorter $SiaT_{spi}$ and $dSiaT_{spi}$-U genes were obtained by PCR to delete the upstream sequence or the 12 amino acid residues in the middle of the $SiaT_{spi}$-U gene, respectively. According to the three dimensional structure of $SiaT_{spi}$, its N-terminus lacking the upstream sequence might be located in the active site. To avoid the loss of activity of $SiaT_{spi}$, the $SiaT_{spi}$ gene was only ligated into pET21a vector to obtain the recombinant $SiaT_{spi}$ with a His-Tag at C-terminus. The $SiaT_{spi}$-U and $dSiaT_{spi}$-U genes were inserted into pET19b adding a His-Tag sequence to the N-terminus of $SiaT_{spi}$-U and $dSiaT_{spi}$-U. The maps of these three recombinant plasmids are shown in Figure 11.3.

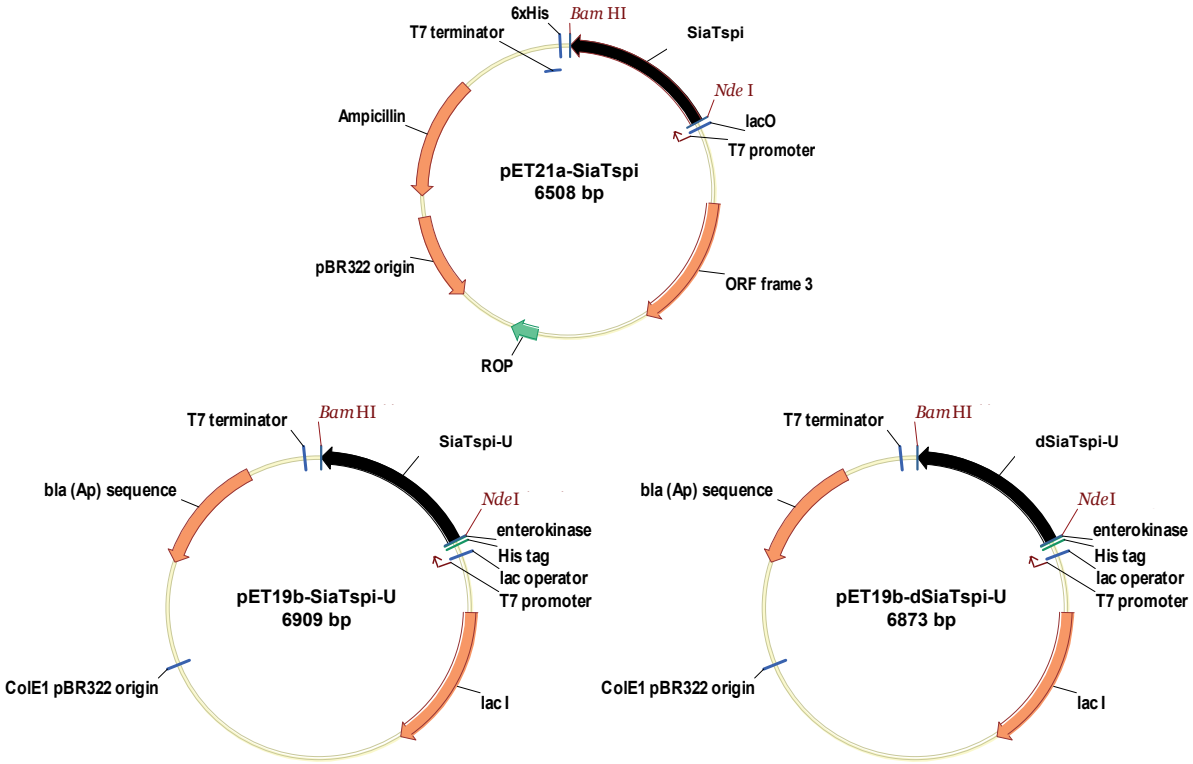



Figure. 11.3. Maps of pET21a-$SiaT_{spi}$, pET19b-$SiaT_{spi}$-U and pET19b-$dSiaT_{spi}$-U

The constructed plasmids were identified by restriction analysis and sequencing, then transformed into BL21 Tuner cells (Novagen) for expression. IPTG was added at different concentration (0.1-1 mM) into the culture during the expression phase to optimize the induction of recombinant proteins at 28°C. The harvested cells were lysed by ultrasonic treatment in an ice bath. Both the supernatant and the pellet of the crude extract were loaded on SDS-PAGE to determine the soluble expression levels. The expression levels within 0.1-1.0 mM concentration of IPTG showed no obvious difference. At 0.1 mM IPTG, the three recombinant $SiaT_{spi}$ could be very well expressed, possibly owed to the codon optimization of $SiaT_{spi}$ gene for *E. coli*. However, unfortunately the expressed proteins were formed as inclusion bodies, and soluble recombinant enzyme was found in the supernatant by SDS-PAGE. The rapid activity determination using Lac-D-T-P-Acr also showed that there was no activity in both crude cell extracts and the supernatant both at 4°C and 25°C. It is surmised that the expression was too fast partly because of the optimization of codon usage of $SiaT_{spi}$ for *E. coli* expression, and that the protein folding occurred in an incorrect fashion which led to inactive $SiaT_{spi}$. Therefore, attempts were started to optimize the expression as described below.

### 11.2.3.Optimization of soluble expression of $SiaT_{spi}$, $SiaT_{spi}$-U and $dSiaT_{spi}$-U in *E. coli*

#### 11.2.3.1. Co-expession with chaperones

The formation of inclusion bodies during the high-level expression of the protein of interest in *E. coli* is usually caused by misfolding of recombinant peptides. *In vivo*, it has been proved that the protein folding is an energy-dependent process mediated by molecular chaperones and foldases.[191] The former bind partially folded proteins and maintain them in a soluble or translocation-competent conformation, while the latter can accelerate the rate-limiting steps along the folding pathway, for example the isomerization of peptidyl-prolyl bonds and the formation and reshuffling of disulfide bridges.[191] As tools useful for protein expression, the chaperonins DnaK-DnaJ-GrpE, GroEL-GroES and Tf in *E. coli* have been well identified and successfully applied for heterologous protein expression, observably promoting the soluble expression level.[192, 193] Therefore, as a first attmpt to improve the soluble expression level in *E. coli* cells, $SiaT_{spi}$, $SiaT_{spi}$-U and $dSiaT_{spi}$-U were coexpressed with combinations of the chaperonins DnaK-DnaJ-GrpE, GroEL-GroES and GroEL-GroES-Tf, respectively. When $SiaT_{spi}$,

SiaT$_{spi}$-U and dSiaT$_{spi}$-U were coexpressed with GroEL-GroES, the soluble expression levels seem to be higher than without chaperonin (Figure 11.4), although the major part of recombinant SiaT$_{spi}$ was still produced in the form of inclusion bodies. To confirm their real activities, fluorescently-labeled lactose was used as acceptor in the presence of CMP-Neu5Ac to detect the potential SiaT activities of both crude extract and supernatant of the cell lysate. The results are shown in Figure 11.5. Without chaperons, SiaT$_{spi}$, SiaT$_{spi}$-U and dSiaT$_{spi}$-U were inactive from both supernatant and crude extract. However, upon coexpression with chaperone GroES-GroEL, all samples of SiaT$_{spi}$, SiaT$_{spi}$-U and dSiaT$_{spi}$-U clearly showed detectable 2,3SiaT activity at romm temperature (20°C) when compared to authentic 2,3- and 2,6-sialated product references. In particular, dSiaT$_{spi}$-U showed much higher 2,3SiaT activity than the others. These results reveal that the SiaT$_{spi}$ indeed has true 2,3SiaT activity and that the chaperons improve the soluble expression level and the correct folding of recombinant SiaT$_{spi}$. On the other hand, the reorganized protein dSiaT$_{spi}$-U showed obviously higher activity than SiaT$_{spi}$ and SiaT$_{spi}$-U. It supports the structure analysis above that the missing upstream sequence is important for the activity of SiaT$_{spi}$. The additional 12 residue sequence, which is suggested to form two additional β-sheets, may give more stability to SiaT$_{spi}$ in a deepsea habitat, but seems not to contribute to stability under normal pressure and temperature. Certainly, the relatice activity of SiaT$_{spi}$, SiaT$_{spi}$-U and dSiaT$_{spi}$-U enzymes ought to be compared in a quantitative way. However, the soluble expression level of SiaT$_{spi}$ was still lower than anticipated, which made the further purification and characterization of SiaT$_{spi}$ impossible. Therefore, other optimization strategies needed be considered.

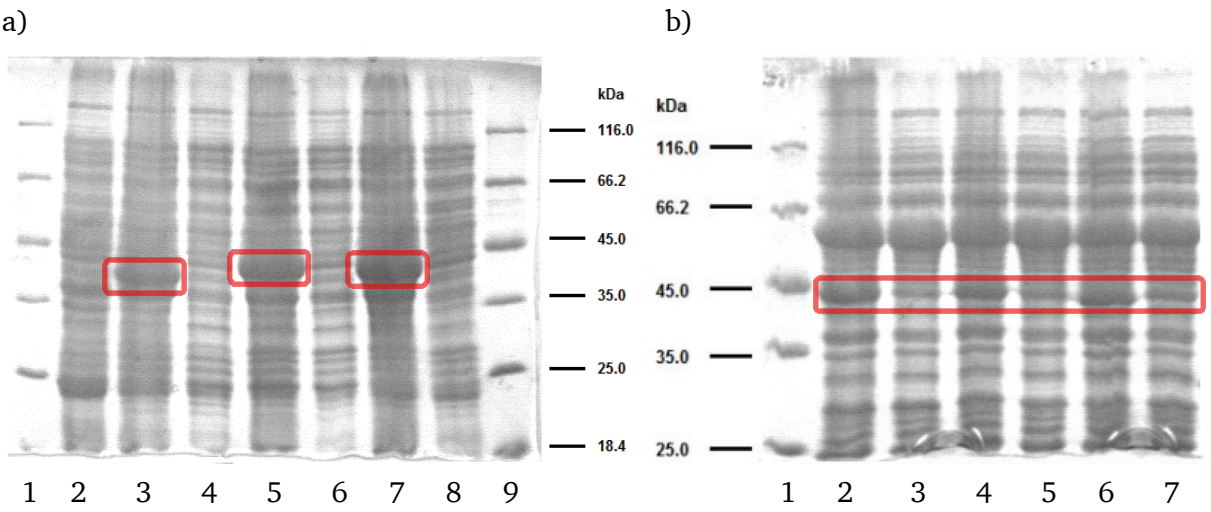



Figure 11.4. Expression level analysis of SiaT$_{spi}$, SiaT$_{spi}$-U and dSiaT$_{spi}$-U. a) Expressed in BL21 Tuner. 1: Marker. 2: Supernatant of BL21 Tuner host cell. 3: Crude cell extract of SiaT$_{spi}$. 4: Supernatant of cell extract of SiaT$_{spi}$. 5: Crude cell extract of SiaTspi-U. 6: Supernatant of cell extract of SiaT$_{spi}$-U. 7: Crude cell extract of dSiaT$_{spi}$-U. 8: Supernatant of cell extract of dSiaT$_{spi}$-U. 9: Marker. b) Coexpression with chaperonins GroEL-GroES. 1: Marker. 2: Crude cell extract of dSiaT$_{spi}$-U. 3: Supernatant of cell extract of dSiaT$_{spi}$-U. 4: Crude cell extract of SiaTspi-U. 5: Supernatant of cell extract of SiaT$_{spi}$-U. 6: Crude cell extract of SiaT$_{spi}$. 7: Supernatant of cell extract of SiaT$_{spi}$.

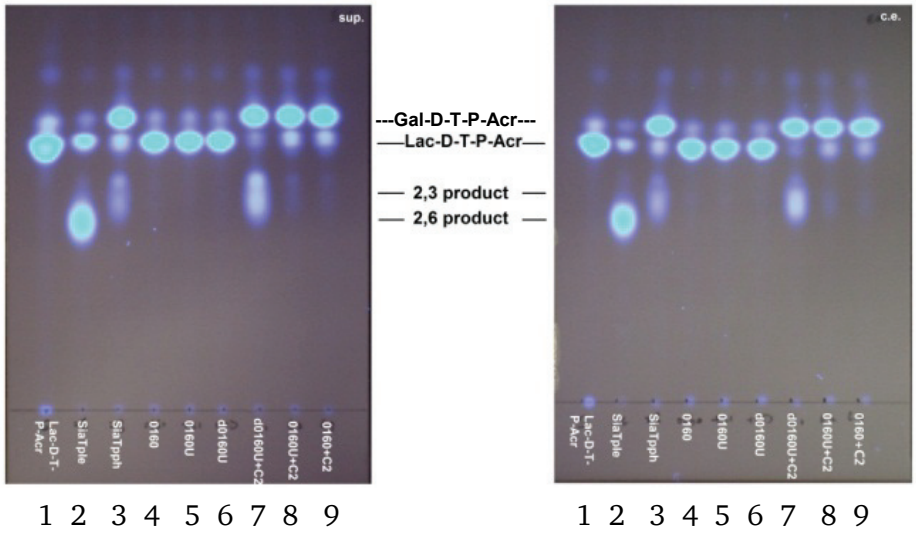



Figure 11.5. Activity determination of $SiaT_{spi}$, $SiaT_{spi}$-U and $dSiaT_{spi}$-U using fluorescent-labeled lactose. In each TLC board from left to right: 1. Lac-D-T-P-Arc, 2. $2,6SiaT_{ple}$, 3. $2,3SiaT_{pph}$, 4. $dSiaT_{spi}U$, 5. $SiaT_{spi}U$, 6. $SiaT_{spi}$, 7. $dSiaT_{spi}U$ coexpressed with GroEL-GroES, 8. $SiaT_{spi}U$ coexpressed with GroEL-GroES, 9. $SiaT_{spi}$ coexpressed with GroEL-GroES. (The competing galactosidase activities shown in 3, 7, 8 and 9 were caused by different host strain BL21(DE3).)

### 11.2.3.2. Expession of $SiaT_{spi}$ at low temperature

*S. piezotolerans* is known to live in a low temperature environment under high pressure, where the enzymes from this microorganism likely have correct folding and optimal activity.[201, 226, 230-233] High pressure is hard to simulate in an average lab because of equipment limitations. Thus, optimization of expression at low temperature became the priority option. On the other hand, a low temperature culture will limit the expression at a relatively low rate, which could also avoid or reduce formation of inclusion bodies. Normal *E. coli* strains can grow at a relative wide range of temperatures from 10°C to 49°C.[234] Therefore, recombinant BL21(DE3) cells carrying both the expression plasmid for $SiaT_{spi}$ and the chaperonins GroEL-GroES were pre-cultured at 37°C until $OD_{600}$ reached 0.5, and then were induced by addition of 0.1 mM IPTG at 16°C for expression. Because of the slow growth of *E. coli* at such a low temperature, the induction was maintained for 2 days in order to harvest sufficient cell mass for subsequent activity determination. The final identification of expression on SDS-PAGE shows that the overall expression level was much lower than that observed at 28°C, but it was pleasing to see that all of the expressed proteins of interest were soluble in the supernatant. As a result, low temperature seemed indeed promising for the soluble expression of recombinant $SiaT_{spi}$. However, no sialyltransfer activity could be detected in the crude extract, which might be due to reduction of the chaperon's activity at low temperature. It has been shown that at 12 °C the activity of the *E. coli* chaperonin GroEL-

GroES retains only about 30% refolding activity, compared to its activity at the temperature optimum of 30°C.[235] Therefore, a special *E. coli* cell line named ArcticExpress DE3 (Agilent, USA) was applied for this particular hypothesis.

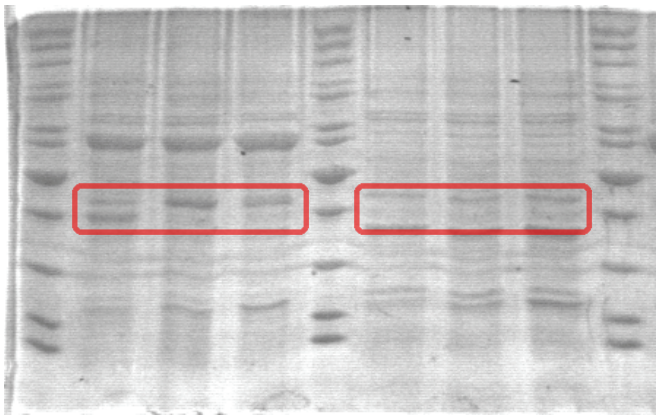



Figure 11.6. Expression level of $SiaT_{spi}$ with chaperonin complex GroEL-GroES or Cpn10-Cpn60 at a low temperature. 1: Marker. 2: Crude cell extract of $SiaT_{spi}$. 3: Crude cell extract of $SiaT_{spi}$-U. 4: Crude cell extract of $dSiaT_{spi}$-U. 5: Marker. 6: Supernatant of cell extract of $SiaT_{spi}$. 7: Supernatant of cell extract of $SiaT_{spi}$-U. 8: Supernatant of cell extract of $dSiaT_{spi}$-U. 9: Marker.

ArcticExpress DE3 is an *E. coli* strain engineered to address the common bacterial gene expression hurdle of protein insolubility. These cells co-express the cold-adapted chaperonins Cpn10 and Cpn60 from the psychrophilic bacterium, *Oleispira Antarctica*, which have high sequence identity to the *E. coli* GroEL and GroES chaperonins (74% and 54%, respectively).[235] Both of proteins show high protein refolding activities even at low temperatures of 4–12°C,[235] improving protein processing and potentially increasing the yield of active protein, which seemed to be perfect for $SiaT_{spi}$ expression. Commercial ArcticExpress DE3 competent cells were transformed with pET21-$SiaT_{spi}$, pET19-$SiaT_{spi}$U and pET19-$dSiaT_{spi}$U plasmids for low-temperature cultivation and protein expression. The recombinant cells were firstly cultured at 37°C overnight, followed by 1 mM IPTG induction for at least 24 h at 12 °C. After that, the cells were harvested by centrifugation at 4°C and soluble expression was determined by SDS-PAGE. The result (Figure 11.6) clearly shows the high soluble expression of the three $SiaT_{spi}$ varieties, similar to the results obtained with chaperonins GroEL-GroES. But it was still disappointing to find that no SiaT activity was detectable in the crude extract. Therefore, the expression of $SiaT_{spi}$ was tried to be carried out in a prokaryotic expression system, *Yarrowia* system, in the next section.

### 11.2.4. Cloning and expression of $SiaT_{spi}$, $SiaT_{spi}$-U and $dSiaT_{spi}$-U in *Yarrowia lipolytica* expression system

*Yarrowia lipolytica* is a GRAS (Generally Recognized as Safe) organism with significant capacities for high molecular weight protein secretion.[236, 237] It also has additional oxidative degradation function for alkanes, fatty acids, fats and oils.[236] Therefore, it has been well used for the preparation of biological intermediates and heterologous gene expression.[236] For its use as an expression host, Barth *et al.* have constructed several vectors containing the regulable *ICL1* promoters (*ICL1* encoding the isocitrate lyase), an rDNA sequence for homologous recombination, a resistance marker for *E. coli* selection, an *ura3* allele for *Y. lipolytica* selection and multiple cloning site for the heterologous gene insert.[236] This vector can insert multiplecopies of the gene of interest into the rDNA range of *Y. lipolytica*'s genome thereby, directly increasing the intracellular expression level of a target protein.

For the expression of $SiaT_{spi}$ in *Y. lipolytica*, a modified p64 vector (named p64PT) was used, which is equipped with all of the features described above (Figure 11.7a and b). The genes for $SiaT_{spi}$-U and $dSiaT_{spi}$-U were inserted into the *Sph*I site of the p64PT vector. For the transformation of the plasmids, the uracil auxotrophic strain H355-S4[238] was used employing a LiAc induced chemical method which has been proved as an efficient way for *Y. lipolytica* recombinant construction.[236] The transformants were selected from minimal medium plates, on which the native host cells cannot survive because of the *ura3* allele deletion. The inserted genes were confirmed by PCR for their presence in *Y. lipolytica* recombinants. (Figure 11.7c) The copy number of the insert gene could be potentially determined by real-time PCR, but in the present case, it was preferred to evaluate its expression and activity first.

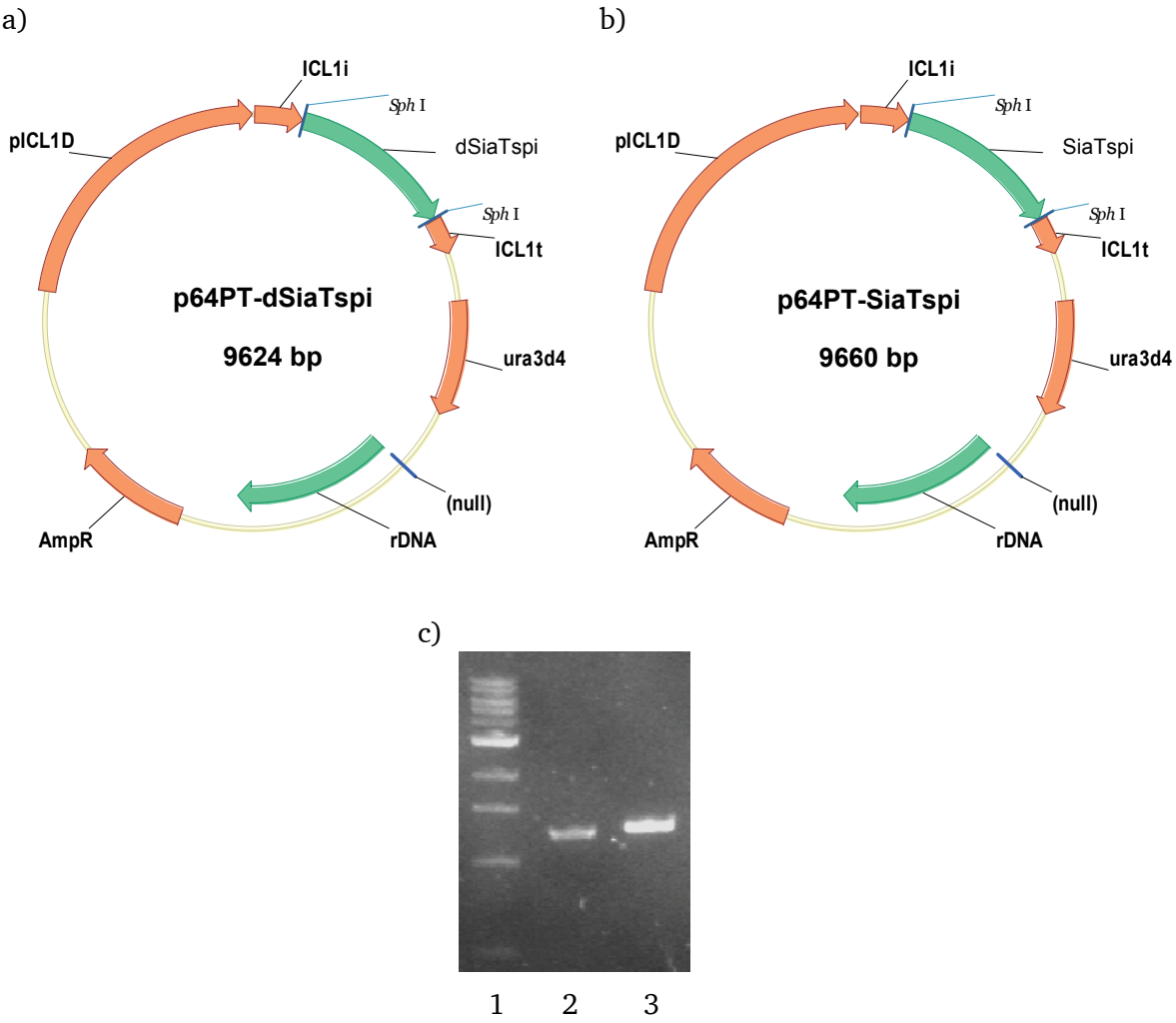



Figure 11.7. Identification of *Y. lipolytica* recombinants by PCR. a) Map of p64P-SiaT-U. b) Map of p64P-dSiaT-U. c) PCR identification of $SiaT_{spi}$ in *Y. lipolytica* recombinants. 1: Marker. 2: $dSiaT_{spi}$-U. 3: $SiaT_{spi}$-U

The fermentation of *Y. lipolytica* was started in minimal medium containing 2% glucose. When glucose was consumed, ethanol was added for induction at 1% final concentration. The $OD_{600}$ of the culture remained stable during the 24 h induction period. For activity tracking, cell samples were taken each 2.5 h. Lysis of cells utilized glass beads with high-speed shaking. For a rapid identification of SiaT activity, Lac-D-T-P-Acr and CMP-Neu5Ac were used as acceptor and donor, respectively. On SDS-PAGE, no target protein band could be detected from all the samples during the induction relative to the host cell H355. The corresponding activity determination showed that all the samples collected during the induction had no SiaT activity at all. This might be due to the very low expression level of target protein in *Yarrowia lipolytica*, partly from the result of the optimized gene codons of $SiaT_{spi}$ for *E. coli* expression

instead of yeast. On the other hand, when considering the much lower number of gene copies in yeast as compared to the multicopy plasmid situation in *E. coli*, the expressed heterologous protein may not be observable by SDS-PAGE from crude extract samples. Thus, the possibility of inactive protein expression cannot be excluded.

## 11.3. Conclusion

$SiaT_{spi}$ is considered a potential SiaT from *Shewanella piezotolerans*. Its sequence BLAST and structure alignment with other 2,3SiaTs in GT80 family retrieved a functional upstream sequence which potentially gives two β-sheets in the 3D-structure of $SiaT_{spi}$. The deletion of a 12 amino acid insert also improved its 2,3SiaT activity when it was co-expressed with chaperonins as demonstrated by a catalyzed reaction using fluorescence labeled acceptor. However, its low soluble expression in the *E. coli* expression system rendered its further property analysis impossible. At low temperatures soluble, $SiaT_{spi}$ could be co-expressed with chaperonins at a low expression level, but activity could not be detected. Its expression was also attempted in the *Y. lipolytica* expression system with a multicopy vector p64PT. However, no obvious $SiaT_{spi}$ production or activity could be detected. Further work may focus on improvement of the expression level first, followed by optimization of active expression. As a potential cold-adapted enzyme, there are difficulties to express $SiaT_{spi}$ well under conventional conditions. Although it was impossible to gain further insight to the properties of $SiaT_{spi}$ except for the activity of its variant $dSiaT_{spi}$-U, the experiments still allowed to estimate the value of methods for enzyme cloning from psychrophiles, for example the application of chaperonins and low temperature cultivation. Further work could be focused on the expression optimization in yeast, including the optimization of the gene sequence and the fermentation conditions. In addition, the use of *Shewanella piezotolerans*'s own chaperonins might also provide a good option for expression of $SiaT_{spi}$ in *E. coli*.

# 12. Summary

## 12.1. Part I. Enzymatic Properties and Directed Evolution of Transketolase

### 12.1.1.Transketolase assay for substrate fingerprinting

Transketolase (TK; EC 2.2.1.1) catalyzes the stereospecific transfer of a two carbon ketol unit to the carbonyl terminus of a variety of aldehydes. The use of TK for carbon-carbon bond formation is receiving increased attention for asymmetric synthesis because of the highly stereocontrolled formation of an (*S*)-configurated chiral center with efficient concomitant chiral resolution for (2*R*)-hydroxyaldehydes. Especially when using hydroxypyruvate (HPA) as donor, the release of $CO_2$ makes the synthetic reaction irreversible, and the released bicarbonate increases the pH of the reaction system. According to this principle, a novel continuous pH-based assay method has been developed for the high-throughput screening of TK, in the presence of HPA as donor and phenol red as a sensitive pH indicator. The assay is generic and can be used for the reliable colorimetric determination of specific enzyme activity and measurement of kinetic constants. Based on this method, the kinetic constants and acceptor specificity of TK from *Escherichia coli* ($TK_{eco}$) and *Saccharomyces cerevisiae* ($TK_{yst}$), as well as two $TK_{eco}$ variants ($TK_{eco}^{D469E}$ and $TK_{eco}^{H26Y}$), have been determined for a detailed comparison of the enzymes. From the quantitative data, the catalytic capabilities of wild-type $TK_{eco}$ and $TK_{sce}$ are quite similar with only very little variation. Wild-type TK enzymes have a strong preference for α-hydroxylated acceptors with strict stereoselectivity towards the (2*R*)-configuration. The variant $TK_{eco}^{D469E}$ shows significantly higher activities towards 2-deoxygenated aldehydes than the corresponding wild-type enzyme.

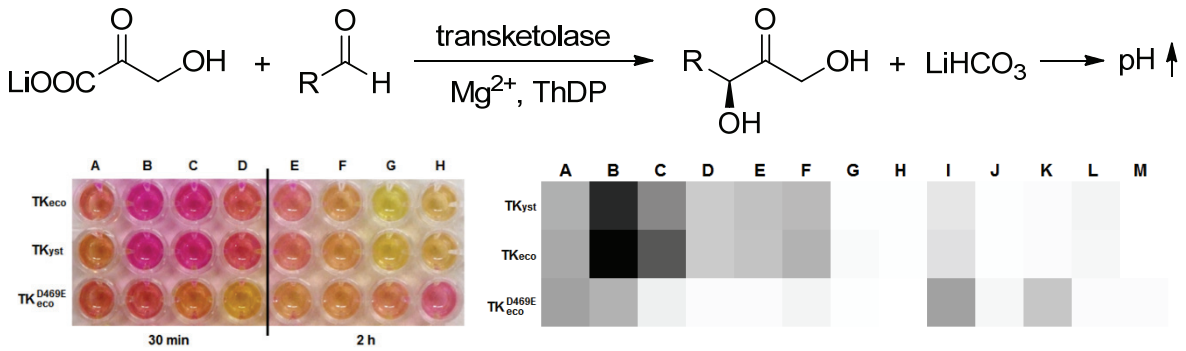



Figure 12.1. The principle of a novel pH-based assay method for transketolase and its application in the determination of specific enzyme activity.

### 12.1.2.Transketolase engineering for generic aldehyde tolerance

The TK from *Geobacillus stearothermophilus* ($TK_{gst}$) is a novel enzyme which has been cloned and identified recently to have an impressively high thermostability. Similar to other TKs, it also shows low activity towards non-α-hydroxylated aldehydes. In order to improve its activity, saturation mutagenesis libraries of $TK_{gst}$ at Leu382 and Asp470 have been constructed in a first attempt to improve activity towards aliphatic aldehydes; Leu382 and Asp470 are the corresponding residues for binding of the hydroxymethylene group of α-hydroxylated aldehydes. Library screening revealed several positive variants from both single-site libraries and the double-site library. Asp469Ile was identified as the top mutant with 17, 13 and 8.6 fold activity improvement towards propanal, butanal and methoxyethanal, respectively, in comparison to the wild-type, but the protein still maintains high thermostability with a $T_m$ of 74 °C.

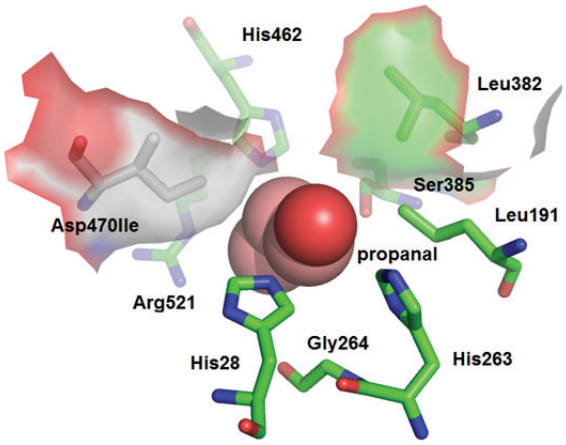



Figure 12.2. Docking of propanal as substrate analog for $TK_{gst}^{D469I}$. Asp470Ile and Leu382 form a non-polar surface to bind the aliphatic chain of propanal with higher affinity.

## 12.2. Part II. Chemoenzymatic Synthesis of neo-Sialoconjugates

### 12.2.1. Sialyltransferase assay for enzyme kinetics

Sialic acids are a family of 9-carbon sugars which exist broadly at the termini of free or conjugated glycans on the surface of cells, playing significant roles in cell physiology. Therefore, it is significant to develop methods for the synthesis of specific sialoconjugate epitopes *in vitro* to facilitate research on carbohydrate-protein interactions in physiological events. α/β-Galactosyl α-2,3/2,6-sialyltransferases transfer sialic acid residues to the terminal galactose moiety of an oligosaccharide, and therefore becomes are useful enzymatic tools to synthesize sialoconjugates *in vitro*. For this purpose, bacterial α-2,3- and 2,6- sialyltransferases have been cloned from *Photobacterium phosphoreum* ($2,3SiaT_{pph}$) and *Photobacterium leiognathi* JT-SHIZ-145 ($2,6SiaT_{ple}$), respectively. In order to understand the substrate specificity of these two SiaTs, a pH-based assay method was established to determine the kinetic constants and substrate tolerance of SiaT. This assay method using a low concentration buffer system, and phenol red as pH indicator, is rapid, sensitive and continuous. Using this method, the kinetic constants and substrate tolerance of $2,3SiaT_{pph}$ and $2,6SiaT_{ple}$ has been determined. These two SiaTs have quite different acceptor specificity. $2,3SiaT_{pph}$ can well accept α- and β-galactosides, lactosides, Glc, Man and GalNAc. However, $2,6SiaT_{ple}$ has high activity only towards β-galactosides and lactosides.

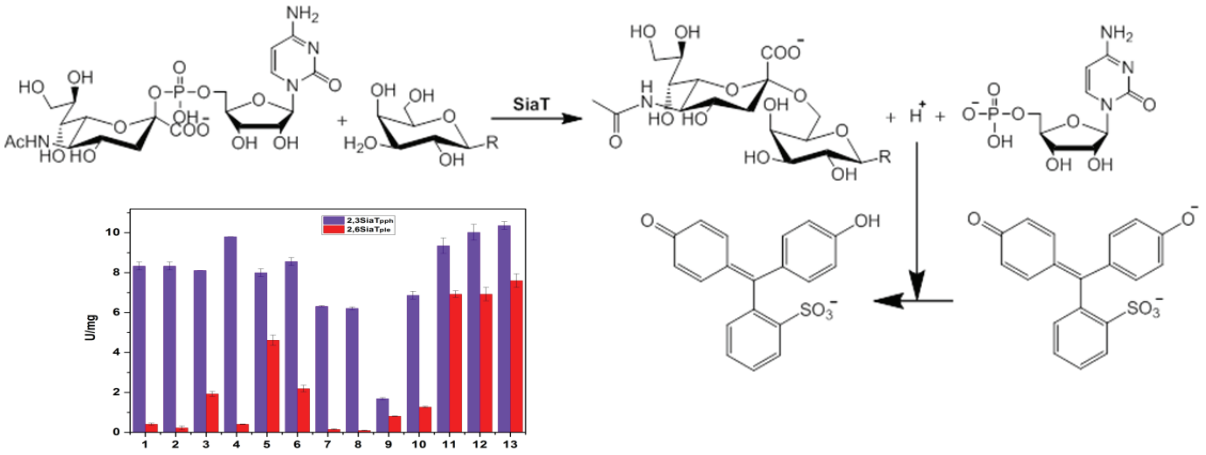



Figure 12.3. The principle of a novel pH-based assay method for sialyltransferase and its application in the determination of specific enzyme activity, with various accepter substrates.

### 12.2.2. Attempted β-SiaT engineering for α-substrate tolerance

From the structure alignment between 2,3SiaT$_{pph}$ and 2,6SiaT$_{ple}$, Trp347 was found to be the responsible residue that is blocking the entrance to the active center for α-galactosides. Therefore, three mutagenesis libraries were constructed to improve 2,6SiaT$_{ple}$'s activity towards α-galactosides using the optimized pH-based assay as high-throughput screening method and methyl-α-D-galactopyranoside as acceptor. However, no positive mutant could be ifentified from the screening, which revealed that Trp347 cannot be simply substituted in 2,6SiaT$_{ple}$. It is speculated that Trp347 contributes to the binding between donor and acceptor and helps to stabilize the correct acceptor conformation to form the 2,6-sialated product.

### 12.2.3. CSS engineering for bulky substrates

CMP-sialic acid synthetase (CSS, E.C. 2.7.7.43) catalyzes the CMP activation of sialic acid using CTP as donor and $Mg^{2+}$ as cofactor. A series of Neu5Ac analogues with bulky modification on the acylamino group were showen to be accepted by CSS from *Neisseria meningitidis* (CSS$_{nme}$), although their corresponding $k_{cat}/K_M$ values were much lower than that of the natural substrate Neu5Ac. From an active site alignment for the acceptor binding of CSS$_{nme}$, Phe192 and Phe193 were identified as the corresponding residues for binding of the acylamino group in Neu5Ac. Therefore, four saturation mutagenesis libraries were

constructed on positions of Phe192 and Phe193 to improve its activity towards bulky analogs Neu5Ac-OBz, Neu5Hex and Neu5Pen$N_3$. From a screening by a pH-based high-throughput screening method, $CSS_{nme}$-F192C/F193Y, $CSS_{nme}$-F192S, and $CSS_{nme}$-F192S/F193Y were the top enzyme variants with 14.1 fold, 4.6 fold and 4.0 fold activity improvements towards Neu5Ac-OBz, Neu5Hex and Neu5Pen$N_3$, respectively. In order to validate the improved activity, the *in situ* CMP activation of the Neu5Ac analogues was applied by using these CSS variants in with 2,6$SiaT_{ple}$ a preparative synthesis of new neo-sialoconjugates. The yields of the final products were 50%, 64% and 70%, repectively.

### 12.2.4. Enzyme cloning for LacNAc synthesis

The *N*-acetyl-D-lactosamine (LacNAc) moiety at the terminus of oligosaccharides is an important natural substrate of SiaT. Synthesis of LacNAc and its derivatives *in vitro* can extend the diversity of acceptors for the synthesis of neo-sialoconjugates. For this purpose, β-1,4-galactosyltransferase from *Helicobacter pylori* ($GalT_{hpy}$) and UDP-galactose 4-epimerase from *E. coli* ($GalE_{eco}$) were cloned and over-expressed in an *E. coli* expression system. Using these two recombinant enzymes, free LacNAc and LacNAc-D-T-P-Acr, as well as Lac-D-T-P-Acr were synthesized from UDP-Glc donor and the corresponding acceptor substrates, with isolated yields of 88%, 92% and 85%, respectively, on a 100 mg scale each for further preparative utilization.

### 12.2.5. Preparative synthesis of neo-sialoconjugates

Using 2,3$SiaT_{pph}$ and 2,6$SiaT_{ple}$, as well as novel $CSS_{nme}$ variants, the neo-2,3- and 2,6-sialoconjugates were synthesized in a one-pot two-enzyme approach. Six typical sialic acids (Neu5Ac, Neu5Gc, KDN, *epi*-KDN, Neu5Ac-9-N3 and Neu5NPhAc) with modifications on C5 and C9 were involved as donors for enzymatic coupling to fluorescent lactose (Lac-D-T-P-Acr) and *N*-acetyl-D-lactosamine (LacNAc-D-T-P-Acr) acceptors with α-2,3- and α-2,6-configuration, respectively. 2,3$SiaT_{pph}$ and 2,6$SiaT_{ple}$ show quite similar donor specificity towards these six sialic acids. Both of them can well accept Neu5Ac, Neu5Gc, KDN and Neu5Ac-9-$N_3$, but are less reactive with the bulky Neu5NPhAc substrate analog. Moreover, 2,6$SiaT_{ple}$ can accept *epi*-KDN better than 2,3$SiaT_{pph}$.

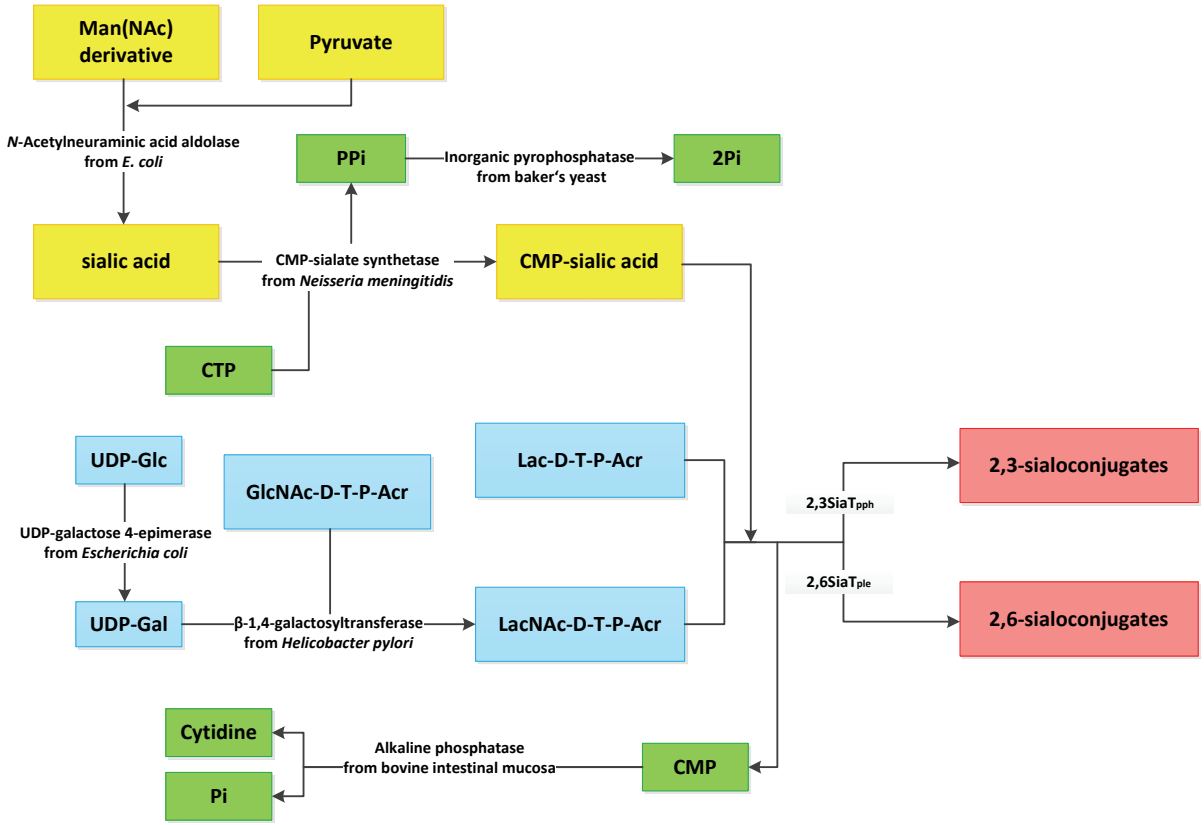



Figure 12.4. Chemoenzymatic synthetic route for the fluorogenic sialoconjugates

### 12.2.6.Characterization of a potential novel SiaT

Recently, a novel potential sialyltransferase from *Shewanella piezotolerans* ($SiaT_{spi}$) has been identified from the genomic *S. piezotolerans* sequence. However, cloning, expression and activity assay of $SiaT_{spi}$ have not been reported yet. Therefore, the artificial gene of this protein was cloned, modified and optimally expressed in the *E. coli* and a *Yarrowia* expression system. Its sequence BLAST and structure alignment with other 2,3SiaTs in CAZy GT80 family retrieved its functional upstream sequence. The deletion of a loop containing 12 amino acid residues also facilitated to identify its 2,3SiaT activity by using a reaction with fluorescent labeled acceptor, particularly when co-expressed with chaperonins. However, its low soluble expression in both *E. coli* and *Y. lipolytica* expression systems prevented its further characterization.

# 13. METHODS

## 13.1. Material, Reagents, Instruments and Services

| Reagent/Instrument | Resource/Formulation | |
|---|---|---|
| **1. Molecular Cloning Kits** | | |
| Wizard Plus SV Minipreps DNA purifcation System | Promega | USA |
| GeneJET Plasmid Miniprep Kit | Fermentas | Germany |
| GeneJET Gel Extraction Kit | Fermentas | Germany |
| Restriction endonucleases | New England Biolab | UK |
| | Fermentas | Germany |
| Phusion High-Fidelity DNA Polymerase | New England Biolab | Germany |
| T4 Ligase | Fermentas | Germany |
| Chaperone Plasmid Set | TaKaRa | Japan |
| QuikChange Lightning Site-Directed Mutagenesis Kit | Agilent | USA |
| QuikChange Lightning Multi Site Directed Mutagenesis Kit | Agilent | USA |
| **2. Biological Reagents** | | |
| LB medium | Tryptone | 10 g |
| | Yeast extract | 5 g |
| | NaCl | 10 g |

| | | |
|---|---|---|
| | $H_2O$ | to 1 L |

*Adjust pH to 7.2, autoclave.

| | | |
|---|---|---|
| Low salt LB medium | Tryptone | 10 g |
| | Yeast extract | 5 g |
| | NaCl | 5 g |
| | $H_2O$ | to 1 L |

*Adjust pH to 7.2, autoclave.

| | | |
|---|---|---|
| Terrific Broth medium | Tryptone | 12g |
| | Yeast extract | 24g |
| | Glycerol | 4 ml |
| | $H_2O$ | to 900 mL |

*Autoclave, cool to 60°C or less before adding 100 ml of filter sterilized 10×TB phosphate (0.17 M $KH_2PO_4$ , 0.72 M $K_2HPO_4$).

| | | |
|---|---|---|
| SOB medium | Tryptone | 20 g |
| | Yeast extract | 5 g |
| | NaCl | 0.5 g |
| | 250 mM KCl | 10 mL |
| | $H_2O$ | to 900 L |

*Adjust pH to 7.0 and add $H_2O$ to 990 mL. Autoclave, cool to room temperature and add 10 mL sterile solution of 1 M $MgCl_2$ before use.

SOC medium

*SOB medium (1 liter) with the addition of 20 mL filter sterilzed 1 M glucose.

| | | |
|---|---|---|
| 10×YNB-Reader | $(NH_4)_2SO_4$ | 30 g |
| | $KH_2PO_4$ | 10 g |
| | $K_2HPO_4 \cdot 3H_2O$ | 1.6 g |

| | | |
|---|---|---|
| | $MgSO_4 \cdot 7H_2O$ | 7 g |
| | NaCl | 5 g |
| | $Ca(NO_3)_2 \cdot 4H_2O$ | 4 g |
| | $H_2O$ | to 1 L |
| *Sterilized by filtration. Store at 4°C. | | |
| Trace-metal stock solution | $H_3BO_3$ | 50 mg |
| | $CuSO_4 \cdot 5H2O$ | 4 mg |
| | KI | 10 mg |
| | $MnSO_4 \cdot 4H_2O$ | 40 mg |
| | $Na_2MoO_4 \cdot 2H2O$ | 20 mg |
| | $ZnSO_4 \cdot 7H2O$ | 40 mg |
| | $H_2O$ | to 100 ml |
| *Sterilized by filtration. | | |
| Iron solution | $FeCl_3$ | 3 g |
| | EtOH | to 100 ml |
| *Sterilized by filtration. Store at 4°C. | | |
| 1000×vitamin solution | Thiaminhydrochlorid | 30 mg |
| | $H_2O$ | to 100 ml |
| *Sterilized by filtration. Store at 4°C. | | |
| Media containing agar (per liter) | agar | 20 g |
| Antibiotics (final concentration) | ampicillin | 100 $\mu$g/mL |
| | chloramphenicol | 37 $\mu$g/mL |
| | kanamycin | 50 $\mu$g/mL |

| | | |
|---|---|---|
| | tetracycline | 12 $\mu$g/mL |
| Galactosides (final concentration) | X-Gal | 20 $\mu$g/mL |
| | IPTG | 0.1 mM |

## 3. Enzymes and Chemical Reagents

The enzymes and chemical reagents used in this thesis were purchased from the following companies: VWR (Germany), Sigma-Aldrich (Germany), Fluka (Germany), Alfa Aesar (Germany), Acros (Germany), Biospherics (USA), Roth (Germany), Riedel-de Haen (Germany) and Applichem (Germany).

## 4. Instruments

| | | |
|---|---|---|
| GeneAmp PCR System 9700 | Applied Biosystems | USA |
| FLUOstar OPTIMA plate reader | BMG Labtech | Germany |
| Biomek Vmax Plate Reader | Beckman | USA |
| Plate Reader | | |
| UV Spectrophotometer UV-1800 | Shimadzu | Japan |
| LKB Ultrospec Plus Spectrophotometer | Pharmacia | Switzerland |
| Biomek 2000 | Beckman | USA |
| TitroLine alpha | | |
| HPLC | Shimadzu | Japan |
| GC-17A/AOC-20i | Shimadzu | Japan |
| Centrifuge 5410, 5415D, 5415R, 5804, 5810 and 5810R | Eppendorf | Germany |
| Biofuge 28 RS | Heraeus | Germany |
| Thermomixer comfort | Eppendorf | Germany |
| Thermomixer 5436 | Eppendorf | Germany |
| Titramax 1000 and Incubator 1000 | Heidolph | Germany |

| | | |
|---|---|---|
| $^{1}$H-NMR and $^{13}$C-HMR | 500 MHz Varian Unity 500 | |
| Step One RT | Applied Biosys | USA |

**5. Services**

| | | |
|---|---|---|
| Gene synthesis | Mr. Gene | Germany |
| Primer synthesis | Biomers | Germany |
| Sequencing | Eurofins MWG | Germany |

## 13.2. Enzymatic Property Study of Transketolase Using pH-Based Assay

### 13.2.1. Cloning of $TK_{eco}$, $TK_{yst}$, and $TK_{eco}^{D469E}$

The gene fragments for wild type $TK_{eco}$ (Genbank Accession No. X68025) and $TK_{yst}$ (Genbank Accession No. X73224) were ligated into pET21a and pET47b vectors, respectively, and transformed into BL21 Tuner or BL21DE3 host strain for the overexpression recombinant TKs. The construction of D469E mutant of TK from *E. coli* ($TK_{eco}^{D469E}$) used QuikChange kit from Agilent (USA). PCR primers 5'-C GGT CTG GGC GAA GAA GGC CCG AC-3' (forward primer) and 5'-G TCG GGC CTT CTT CGC CCA GAC CG-3' (backward primer) were designed for this certain site-directed mutation.

PCR reaction:

| | |
|---|---|
| dsDNA template: | 2 $\mu$l |
| Primer 1: | 2 $\mu$l |
| Primer 2: | 2 $\mu$l |
| dNTP mix: | 1 $\mu$l |
| 10×Reaction Buffer: | 5 $\mu$l |
| QuikSolution reagent: | 1.5 $\mu$l |
| QuikChange Lightning Enzyme: | 1 $\mu$l |
| $ddH_2O$: | up to 50 $\mu$l |

| | | |
|---|---|---|
| 95°C | 2 min | |
| 95°C | 30 sec | |
| 60°C | 10 sec | |
| 68°C | 4 min | 30 cycles |
| 68°C | 10 min | |

The PCR product was identified by electrophoresis and digested by 2 $\mu$l *Dpn*I for 5 min at 37°C. Then 2 $\mu$l PCR products were transformed into Jm109 competent cells for plasmid amplification and cultured on Amp-agar plate overnight. The positive colonies on Amp-agar plate were picked and cultured in 5 mL Amp-LB medium overnight. The amplified plasmid was then extracted and sent for sequencing. After that, the positive mutant plasmid was transformed into BL21 Tuner competent cells (Novagen, Germany) for protein expression.

### 13.2.2.Expression and purification of TKs

All of the recombinant cells were cultured in a 10 L fermenter using Amp- or Kan-TB medium (dependent on the resistance of its vector) at 37°C until $OD_{600}$ reached 0.5. Then the cells were induced by 0.1 mM IPTG at 28°C for the expression. After 6 h, the cells were harvested by centrifugation and stored at -80°C. For the purification, 20 g cells were lysed by 0.5 mg/mL lysozyme (Roth, Germany) in 200 mL 20 mM Tris (pH 7.5) containing 0.5 M NaCl, and ultrasonic to obtain crude extract of the cells. After 12,000 rpm centrifugation for 30 min, the supernatant (added 25 mM imidazole) was loaded on Ni- affinity columns (GE, USA) to purify the recombinant protein. After sample loading, the affinity columns were washed by 50 mM imidazole buffer. Then, the target protein was washed out by 200 mM imidazole buffer. The buffer of purified TK was replaced by 2 mM TEA buffer (pH 7.5) using ultrafiltration. The purified enzymes were stored in 50% glycerol at -20°C.

### 13.2.3.Development of a pH-based assay method for TK

#### 13.2.3.1. Calibration curve, limit of detection (LOD) and limit of quantification (LOQ)

The standard curve of the assay method used a series of finial concentrations of $NaHCO_3$ from 0 mM to 1.0 mM in a 96-well plate. The mixture also contained 50 mM LiHPA, 2.4 mM ThDP, 9 mM $MgCl_2$, 40 $\mu$l $TK_{yst}$ (0.3 mg/mL), 2 $\mu$l 0.1% phenol red and 2 mM TEA (pH 7.5). Total volume was 200 $\mu$l. The relationship between concentration of $HCO_3^-$ and absorbance was measured at 560 nm by Vmax plate reader (Beckman, USA) to obtain the standard curve $y = ax + b$. LOD was defined as $LOD = 3.3Sb / \bar{a}$, and LOQ was defined as $LOD = 10Sb / \bar{a}$, where $Sb$ is the standard deviation of the control.

#### 13.2.3.2. Reaction system for the assay

The assay reaction was carried out on 96-well flat bottom plate. The reaction mixture contained 40 $\mu$l TK (0.1, 0.3 or 0.8 mg/mL, dependent on the activity of TK), 50 mM or 200

mM acceptor (dependent on the activity of TK towards this acceptor), 2.4 mM ThDP, 9 mM $Mg^{2+}$, 2 mM TEA (pH 7.5) and 2 $\mu$l phenol red (0.1% , 2.82 mM). 50 mM LiHPA was added to start the reaction. The total volume in each well was 200 $\mu$l. The absorbance increase was measured at 560 nm by plate reader at 20°C.

#### 13.2.3.3. Dependence of rate on amount of TK

The specific activity measurement at varying enzyme concentrations used a series of amount of 0.3 mg/mL $TK_{yst}$ (from 0 $\mu$l to 40 $\mu$l). 50 mM D-glyceraldehyde was used as the acceptor. Absorbance data for the first 60 sec was recorded and converted into concentration of $HCO_3^-$ according to the calibration curve for the activity calculation.

### 13.2.4. Sbustrate tolerance screening and kinetic study of TKs

#### 13.2.4.1. Determination of acceptor specificity

To screening the acceptor tolerance of TKs, the activities towards various acceptors were measured, including 50 mM formaldehyde, glycolaldehyde, D,L-glyceraldehyde, D-glyceraldehyde, D-erythrose, L-threose and D,L-2,4-dihydroxy butanal, and 200 mM D-ribose, L-lyxose, D-xylose, L-arabinose, acetaldehyde, 3-hydroxpropanl, propionaldehyde, (*R*)-3,4-dihydroxy butanal, D-2-deoxy-ribose, D-allose, D-gulose, L-gulose, L-mannose, and D-glucose. Suitable concentrations of TKs were added into the reaction according to the activities toward different acceptors (0.1 mg/mL for formaldehyde, 0.3 mg/mL for glycolaldehyde, D,L-glyceraldehyde, D-glyceraldehyde, D-erythrose, L-threose and D,L-2,4-dihydroxy butanal, and 0.8 mg/mL for D-ribose, L-lyxose, D-xylose, L-arabinose, acetaldehyde, 3-hydroxpropanl, propionaldehyde, (*R*)-3,4-dihydroxy butanal, 2-deoxy-D-ribose, D-allose, D-gulose, L-gulose, L-mannose and D-glucose).

### 13.2.4.2. Determination of kinetic constants

The enzymatic kinetics towards both donor LiHPA and acceptor (glycolaldehyde, D,L-glyceraldehyde, D-glyceraldehyde and D-erythrose) were determined at varying concentration of substrates. The reaction rate at each concentration was calculated according to the calculation curve and collected for non-linear fit to the Michaelis-Menten equation using Origin 8 software to obtain enzymatic kinetic values.

## 13.3. Directed Evolution of Transketolase from *Geobacillus stearothermophilus*

### 13.3.1.1. Building of Leu382 and Asp470 single-site mutagenesis libraries of $TK_{gst}$

For the construction of Leu382 and Asp470 single-site mutagenesis libraries of $TK_{gst}$, the QuikChange kit from Agilent (USA) was used. PCR primers TKGSL382f 5'- TTT GGC GGT TCG GCG GAC NNS GCA AGC TCG AAT AAA ACG C-3' (forward primer) and TKGSL382b 5'-G CGT TTT ATT CGA GCT TGC SNN GTC CGC CGA ACC GCC AAA-3' (backward primer) and TKGSD470f 5'-C AGC ATC GCC GTC GGC GAA NNS GGG CCG ACG CAC-3' (forward primer) and TKGSD470b 5'-GTG CGT CGG CCC SNN TTC GCC GAC GGC GAT GCT G-3' (backward primer) were designed for Leu382 and Asp470 sites, respectively. The PCR used the artificial $TK_{gst}$ gene in vector pET47b (pET47b-$TK_{gst}$) as template.

PCR reaction for Leu382:

PCR reaction:

| | |
|---|---|
| dsDNA template: | 2 $\mu$l |
| Primer 1: | 2 $\mu$l |
| Primer 2: | 2 $\mu$l |
| dNTP mix: | 1 $\mu$l |
| 10×Reaction Buffer: | 5 $\mu$l |
| QuikSolution reagent: | 1.5 $\mu$l |
| QuikChange Lightning Enzyme: | 1 $\mu$l |
| $ddH_2O$: | up to 50 $\mu$l |

| | | |
|---|---|---|
| 95°C | 2 min | |
| 95°C | 30 sec | |
| 72°C | 10 sec | |
| 68°C | 4 min | 30 cycles |
| 68°C | 10 min | |

PCR reaction for Asp470:

PCR reaction:

| | |
|---|---|
| dsDNA template: | 2 $\mu$l |

| | |
|---|---|
| Primer 1: | 2 $\mu$l |
| Primer 2: | 2 $\mu$l |
| dNTP mix: | 1 $\mu$l |
| 10×Reaction Buffer: | 5 $\mu$l |
| QuikSolution reagent: | 1.5 $\mu$l |
| QuikChange Lightning Enzyme: | 1 $\mu$l |
| dd$H_2O$: | up to 50 $\mu$l |

| | | |
|---|---|---|
| 95°C | 2 min | |
| 95°C | 30 sec | |
| 68°C | 10 sec | |
| 68°C | 4 min | 30 cycles |
| 68°C | 10 min | |

The PCR products were digested by 2 $\mu$l *Dpn*I for 1 h at 37°C. Then 2 $\mu$l PCR products was transformed into BL21(DE3) competent cells and cultured on kanamycin agar plates overnight. 192 colonies were picked into 96-well plates containing 150 $\mu$l/well LB-kanamycin medium. After overnight culture, 30 $\mu$l glycerol was added into each well. The whole plates were sealed with plastic lids and stored at -80°C.

#### 13.3.1.2. Building of Leu382-Asp470 double-site mutagenesis library of $TK_{gst}$

For the construction of the Leu382-Asp470 double-site mutagenesis libraries of $TK_{gst}$, the QuikChange kit from Agilent (USA) was used. PCR primers TKGSL382f 5'- TTT GGC GGT TCG GCG GAC NNS GCA AGC TCG AAT AAA ACG C-3' (forward primer) and TKGSL382b 5'-G CGT TTT ATT CGA GCT TGC SNN GTC CGC CGA ACC GCC AAA-3’ (backward primer) and TKGSD470f 5'-C AGC ATC GCC GTC GGC GAA NNS GGG CCG ACG CAC-3' (forward primer) and TKGSD470b 5'-GTG CGT CGG CCC SNN TTC GCC GAC GGC GAT GCT G-3’ (backward primer) were designed for Leu382 and Asp470 sites, respectively. The PCR used the plasmid mixture of the Asp470 library as template.

PCR reaction:

| | |
|---|---|
| dsDNA template: | 2 $\mu$l |
| Primer 1: | 2 $\mu$l |
| Primer 2: | 2 $\mu$l |

| | |
|---|---|
| dNTP mix: | 1 $\mu$l |
| 10×Reaction Buffer: | 5 $\mu$l |
| QuikSolution reagent: | 1.5 $\mu$l |
| QuikChange Lightning Enzyme: | 1 $\mu$l |
| dd$H_2O$: | up to 50 $\mu$l |

| | | |
|---|---|---|
| 95°C | 3 min | |
| 95°C | 30 sec | |
| 72°C | 10 sec | |
| 68°C | 4 min | 30 cycles |
| 68°C | 10 min | |

The PCR product was digested by 2 $\mu$l *Dpn*I for 1 h at 37°C. Then 2 $\mu$l PCR products was transformed into BL21(DE3) competent cells and cultured on kanamycin agar plates overnight. 3,456 colonies were picked into 96-well plates containing 150 $\mu$l/well LB-kanamycin medium. After overnight culture, 30 $\mu$l glycerol was added into each well. The whole plates were sealed with plastic lids and stored at -80°C.

#### 13.3.1.3. Development of screening method for mutagenesis library

$TK_{eco}$ and $TK_{yst}$ recombinant cells were used as samples for the development of a screening method for the mutagenesis library. BL21(DE3) host cell was used as control. 2 $\mu$l of each clone was transferred to individual wells in 96-well plates containing 150 $\mu$l/well LB growth media comprised with 30 $\mu$g/mL kanamycin. The plates were sealed with plastic lids to avoid evaporation, and then incubated at 30°C, 900 rpm overnight. After that, 5 $\mu$l was transferred from each well to 96-deepwell plates with plastic lids containing 400 $\mu$l/well of LB-kanamycin growth medium with 0.1 mM IPTG. The plates were then incubated at 30°C, 900 rpm overnight. Cells were harvested by centrifugation at 4000 rpm for 30min. The culture medium was removed and the cell pellets could be stored at -80°C.

The cell pellets were resuspended in 150 $\mu$l resuspension buffer (1/10 BugBuster solution (Novagen, Germany), 0.5 mg/mL lysozyme (Roth, Germany) and 4U/mL Benzonase endonuclease (Novagen, Germany)) and incubated in a shaker for 0.5 to 1 hour at room temperature. Then, the cell lysate was centrifugated at 4000 rpm for 30min. 40 $\mu$l of the supernatant was transferred to a new 96-well plate. 140 $\mu$l of substrate solution (2 mM TEA

(pH7.5), 9 mM $MgCl_2$, 2.4 mM ThDP, 2.82 mM phenol red and 2.5 mM (for $TK_{eco}$) or 5 mM (for $TK_{yst}$) glycolaldehyde) was added to each well. The reaction was initiated by the addition of 50 mM LiHPA. The OD increase was read by the plate-reader at 560 nm.

For the evaluation and validation of this assay method, *Z*-factor was used as calculated using the data of the acceptor specificity measurement according to the formula $Z = 1 - (3\sigma_s + 3\sigma_c) / |\mu_s + \mu_c|$, where $\sigma_s$ is sample standard deviation, $\sigma_c$ is control standard deviation, $\mu_s$ is sample average, and $\mu_c$ is control average. 48 samples (half 96-well plate) of each TK and host cells were used for *Z*-factor calculation.

### 13.3.1.4. Screening of mutagenesis libraries of $TK_{gst}$

2 $\mu$l of each clone in libraries was transferred to individual wells in 96-well round bottom plates containing 150 $\mu$l/well LB growth media comprised 30 $\mu$g/mL kanamycin. The plates were sealed with plastic lids to avoid evaporation, and then incubated at 30°C, 1,000 rpm overnight. After that, 5 $\mu$l was transferred from each well to 96-deepwell plates with plastic lids containing 400 $\mu$l/well of LB-kanamycin growth medium containing 0.1 mM IPTG. The plates were then incubated at 30°C, 1,000 rpm overnight, until the amount of cells in each well reached the maximum. Cells were harvested by centrifugation at 4,000 rpm for 20 min. The culture medium was removed and the cell pellets could be stored at -80°C.

The cell pellets were resuspended in 150 $\mu$l resuspension buffer (1/10 BugBuster solution (Novagen, Germany), 0.5 mg/mL lysozyme (Roth, Germany) and 4U/mL Benzonase endonuclease (Novagen, Germany)) and incubated in a shaker for 30 min at room temperature. Then, the cell lysate was centrifugated at 4,000 rpm for 30 min. 40 $\mu$l of the supernatant was transferred to a new 96-well flat bottom plate. 140 $\mu$l of substrate solution (2 mM TEA (pH7.5), 9 mM $MgCl_2$, 2.4 mM ThDP, 2.82 mM phenol red and 200 mM propionaldehyde or 100 mM L-glyceraldehyde was added to each well. The reaction was initiated by the addition of 50 mM LiHPA. The total volume in each well was 200 $\mu$l. The OD increase was read by the plate-reader at 560 nm.

20 positive colonies with the highest activity were picked from each plate into a new plate for the second round screening and so on.

### 13.3.1.5. Expression and purification of positive mutants

10 $\mu$l from each positive mutant was transferred into 1 mL Kan-LB medium and cultured at 37°C 250 rpm overnight. Then, 100 $\mu$l each culture was transferred into 20 mL Kan-LB medium and cultured at 37°C 250 rpm until OD reached 0.5. 0.1 mM IPTG (final concentration) was added for the expression induction at 28°C overnight. After that, the cells were harvested by centrifugation at 4,000 rpm for 20 min and stored at -80°C.

For the purification, the harvested cells were lysed by 3 mL lysis buffer which contains 0.5 mg/mL lysozyme, 4 U/mL benzonase endonuclease, 1/10 Bugbuster solution, 10% glycerol, 2.4 mM ThDP, and 9 mM $MgCl_2$). After the incubation at room temperature for half hour, the lysates were incubated at 50 °C for half hour. Then the lysates were treated with 4,000 rpm centrifugation for 30 min. The supernatants were used as purified enzyme samples. The purity of these samples was identified by SDS-PAGE. Their concentration was measured by Brandford method.

### 13.3.1.6. Identification of positive mutants

The assay reaction was carried out on 96-well flat bottom plate. The reaction mixture contained 0.04 $\mu$g TK, 200 mM propanal, methoxypropanal or buthanal, 2.4 mM ThDP, 9 mM $Mg^{2+}$, 2 mM TEA (pH 7.5) and 2 $\mu$l phenol red (0.1% , 2.82 mM). 50 mM LiHPA was added to start the reaction. The total volume in each well was 200 $\mu$l. The absorbance increase was measured at 560 nm by plate reader at 20°C.

### 13.3.1.7. Determination of thermostability for positive mutants

20 $\mu$l 0.5 mg/mL TK samples and 5 µl Sypro Orange (1:500) (Sigma) were mixed in a PCR plate. The real-time PCR program was 25°C-95°C, 1°C per min. The fluorescence data was recorded and processed by real time PCR software.

## 13.4. Chemoenzymatic Synthesis of *N*-Acetyl-D-Lactosamine Precursors

### 13.4.1. Cloning, expression and characterization of β-1,4-galactosyltransferase from *Helicobacter pylori*

#### 13.4.1.1. Cloning of β-1,4-galactosyltransferase from *Helicobacter pylori*

The DNA sequence of the $GalT_{hpy}$ gene from *Helicobacter pylori* was optimized for expression in *E. coli* and synthesized by the company Mr.Gene. *Nde*I and *Bam*HI restriction enzyme sites were added at 5'- and 3'- terminuses, respectively, for further cloning work. The plasmid pMA-HPGT containing the synthesized $GalT_{hpy}$ gene from the company Mr.Gene was transformed into *E. coli* Jm109 competent cells for plasmid amplification. The recombinant cells were cultured in 5 mL LB medium comprising 50 $\mu$g/mL ampicillin with 250 rpm shaking at 37°C overnight. The plasmid was extracted and digested by *Nde*I and *Bam*HI, and then purified by gel extraction. After that, the $GalT_{hpy}$ gene fragment was ligated into pET19b vector to construct the expression plasmid pET19b-$GalT_{hpy}$, which was then transformed into *E. coli* Jm109 competent cells for plasmid amplification and identification. The positive colonies were picked and cultured in 5 mL Amp-LB medium at 37°C with 250 rpm shaking overnight. The extracted plasmids were identified by *Nde*I and *Bam*HI restriction endonuclease enzyme digestion and electrophoresis, as well as sequencing. Then, the identified final plasmid was transformed into *E. coli* BL21 Tuner competent cells. The positive colony (BL21T-pET19b-$GalT_{hpy}$) was picked and cultured in 5 mL Amp-LB medium at 37°C with 250 rpm shaking overnight and stored in 16% glycerol at -80°C.

```
1       CATATGCTGC GCGTGTTTAT CATTAGCCTG AACCAGAAAG TTTGTGACAC
        CTTCGGACTG GTATTCCGTG ATACTACCAC CCTGCTGAAT AACATTAACG
101     CCACACACCA CCAGGCACAG ATCTTTGATG CCATTTATAG CAAAACGTTT
        GAGGGCGGAC TGCACCCACT GGTCAAAAAA CATCTGCACC CGTATTTCAT
201     TACCCAGAAC ATTAAAGACA TGGGCATTAC CACCAACCTG ATCAGCGAAG
        TGAGCAAATT CTATTATGCT CTGAAATATC ACGCCAAATT CATGTCACTG
301     GGCGAACTGG GTTGTTATGC CAGCCATTAT TCCCTGTGGG AGAAATGTAT
        CGAGCTGAAC GAAGCCATTT GTATCCTGGA GGACGACATT ACACTGAAAG
```

```
401     AGGACTTTAA AGAGGGCCTG GATTTCCTGG AAAAACACAT TCAGGAGCTG
        GGCTATGTTC GTCTGATGCA TCTGCTGTAT GACCCTAACG TGAAAAGCGA
501     GCCTCTGAAC CATAAAAATC ATGAGATTCA GGAACGTGTC GGCATTATCA
        AAGCCTATAG CCATGGTGTT GGTACTCAGG GTTATGTGAT TACGCCGAAA
601     ATTGCCAAAG TGTTCAAAAA ACATTCCCGT AAATGGGTGG TTCCGGTTGA
        TACGATCATG GATGCCACCT TTATCCACGG GGTGAAAAAT CTGGTACTGC
701     AGCCGTTTGT GATCGCCGAT GATGAGCAAA TCAGCACCAT TGCCCGTAAA
        GAAGAGCCGT ATTCCCCAAA AATCGCTCTG ATGCGCGAAC TGCACTTCAA
801     ATATCTGAAA TATTGGCAGT TTGTGTGAGG ATCC
```

Sequence 13.1. Artificial gene sequence of $GalT_{hpy}$

### 13.4.1.2. Expression of β-1,4-galactosyltransferase from *Helicobacter pylori*

10 $\mu$l BL21T-pET19b-$GalT_{hpy}$ was transferred into 250 mL LB medium and cultured at 37°C with 250 rpm shaking overnight. Then the 250 mL culture was transferred into 5 L TB medium and cultured in a 10 L fermenter at 37°C with 500 rpm stir until $OD_{600}$ reached 0.5. 0.1 mM IPTG was added into the culture for the expression induction at 28°C for at least 6 hours. After that, the cells were harvested by centrifugation at 4,000 rpm for 20 min. The well cells were frozen and stored at -80°C. The recombinant $GalT_{hpy}$ expression level was identified by SDS-PAGE.

### 13.4.1.3. Rapid characterization of β-1,4-galactosyltransferase from *Helicobacter pylori*

The rapid identification of recombinant $GalT_{hpy}$ used a 100 $\mu$l reaction, in which it contained 3 mM Glc-D-T-P-Acr as acceptor, 5 mM UDP-Gal as donor, 20 mM Tris (pH8.0) as buffer and 10 mM $Mn^{2+}$ as cofactor. 50 $\mu$l supernatant of cell lysate of BL21T-pET19b-$GalT_{hpy}$ was added as catalyst to start the reaction. After 30 min, the Lac-D-T-P-Acr product was identified on TLC plate with mobile phase n-BuOH:aceton:AcOH:$H_2O$=40:40:7:13 .

### 13.4.1.4. Coexpression of β-1,4-galactosyltransferase from *Helicobacter pylori* with chaperones

The chaperone kit Chaperone Plasmid Set was purchased from Takara (Japan). The chaperone plasmids pG-KJE8, pGro7, pKJE7, pG-Tf2 and pTf16 were transformed into BL21(DE3) competent cells and cultured on LB-agar plates with chloramphenicol. The positive colonies were picked and cultured in LB medium with chloramphenicol at 37°C overnight. Then the recombinants were stored in 16% glycerol at -80°C.

Chaperone BL21(DE3) cells was cultured in 5 ml LB medium with chloramphenicol at 37°C with 250 rpm shaking for 16-18 h. 2.5 ml each culture was transferred into 250 ml SOB medium and cultured at 37°C with 250 rpm shaking until $OD_{600}$ reaches 0.6-0.8. Then the cultures were put in ice for 10 min. After that, the cells were harvested by 2,000g centrifugation at 4°C for 10 min, followed by adding 80 ml cold TB medium to resuspend the cells. The cells were put in ice for 10 min again and centrifuged the resuspension at 2,000g 4°C for 10 min to remove the TB medium. Then 20ml TB medium was added and the cells were mixed gently. The resuspension was treated by ice for 10 min. Finally, 1.5 ml DMSO and 75 mM DTT were added and mixed gently. The cells were frozen by liquid nitrogen and stored at -80°C.

The pET19b-$GalT_{hpy}$ plasmid was transformed into competent chaperone cells and cultured on LB agar plate with ampicillin and chloramphenicol at 37°C overnight. The positive colonies were picked and identified by plasmid identification. After that, these five recombinant cells were stored in 16% glycerol and stored at -80°C.

### 13.4.1.5. Gene modification of β-1,4-galactosyltransferase from *Helicobacter pylori*

The pET19b-$GalT_{hpy}$ was prepared as template for the PCR of the gene modification. Primers d27GalT19-f 5'- TGA GGA TCC GGC TGC TAA CAA AG -3' and d27GalT19-b 5'- GGC AAT GGT GCT GAT TTG CTC AT -3', and d27GalT21-f 5'- GGA TCC GAA TTC GAG CTC CGT -3' and d27GalT21-b 5'- GGC AAT GGT GCT GAT TTG CTC AT -3' were designed synthesized by Biomers (Germany). The PCR reaction used Phusion High-Fidelity DNA Polymerase

(Fermentas, Germany).

PCR reaction:

| | |
|---|---|
| dsDNA template: | 2 $\mu$l |
| Primer 1: | 2 $\mu$l |
| Primer 2: | 2 $\mu$l |
| dNTP mix: | 1 $\mu$l |
| 10×Reaction Buffer: | 5 $\mu$l |
| DNA Polymerase: | 0.5 $\mu$l |
| dd$H_2O$: | up to 50 $\mu$l |

| | | |
|---|---|---|
| 97°C | 2 min | |
| 97°C | 10 sec | |
| 72°C | 10 sec | |
| 72°C | 4 min | 30 cycles |
| 72°C | 10 min | |

The PCR products were identified by electrophoresis and purified by gel extraction. Then, they were ligated into pUC19 vector for sub cloning. The ligated products were transformed into Jm109 competent cells and cultured on Amp-LB-agar plates at 37°C overnight. The positive colonies were picked and cultured in 5 mL amp-LB medium at 37°C with 250 rpm shaking overnight. Then their plasmids were extracted and identified by *Nde*I and *Bam*HI restriction endonuclease enzyme digestion and electrophoresis. The positive plasmid was digested by *Nde*I and *Bam*HI at 37°C overnight. The $dGalT_{phy}$ gene fragments were purified by gel extraction and then ligated into pET19b and pET21a vector to obtain pET19b-$dGalT_{hpy}$ and pET21a-$dGalT_{hpy}$ plasmid, respectively. The ligation products were transformed into Jm109 competent cells and cultured on Amp-LB-agar plates at 37°C overnight. The positive colonies were picked and cultured in 5 mL Amp-LB medium at 37°C with 250 rpm shaking overnight. Then their plasmids were extracted and identified by *Nde*I and *Bam*HI restriction endonuclease enzyme digestion and electrophoresis, as well as sequencing. The positive plasmid was then transformed into BL21 Tuner competent cells and cultured on Amp-LB-agar plates at 37°C overnight. The positive colonies were picked and cultured in 5 mL amp-LB medium at 37°C with 250 rpm shaking overnight, and then stored in 16% glycerol at -80°C.

10 $\mu$l BL21T-pET19b-$dGalT_{hpy}$ and BL21T-pET21a-$dGalT_{hpy}$ were transferred into 5 mL LB medium, respectively, and cultured at 37°C with 250 rpm shaking until OD reached 0.5. 0.1 mM IPTG was added into the culture for the expression induction at 28°C with shaking

overnight. After that, the cells were harvested by centrifugation at 4,000 rpm for 20 min. The recombinant $dGalT_{hpy}$ expression levels were identified by SDS-PAGE. The identification of recombinant $dGalT_{hpy}$ activity was used the method described in 13.4.1.3.

## 13.4.2. Cloning, expression and characterization of UDP-galactose 4-epimerase from *Escherichia coli*

### 13.4.2.1. Cloning of UDP-galactose 4-epimerase from *Escherichia coli*

The genome DNA of *E. coli* K12 was prepared as template for the PCR of $GalE_{eco}$ cloning. Primers GalEf (forward primer) 5'-GGA ATT CCA TAT GAG AGT TCT GGT TAC CGG TGG-3' and GalEb (backward primer) 5'-CGC GGA TCC TTA ATC GGG ATA TCC CTG T-3' were designed synthesized by Biomers (Germany).

PCR reaction:

| | |
|---|---|
| dsDNA template: | 2 $\mu$l |
| Primer 1: | 2 $\mu$l |
| Primer 2: | 2 $\mu$l |
| dNTP mix: | 1 $\mu$l |
| 10×Reaction Buffer: | 5 $\mu$l |
| DNA Polymerase: | 0.5 $\mu$l |
| $ddH_2O$: | up to 50 $\mu$l |

| | | |
|---|---|---|
| 97°C | 2 min | |
| 97°C | 10 sec | |
| 72°C | 3 min | 30 cycles |
| 72°C | 10 min | |

The PCR product was identified by electrophoresis and purified by gel extraction. Then, it was ligated into pUC19 vector for sub cloning. The ligated product was transformed into Jm109 competent cells and cultured on Amp-LB-agar plate at 37°C overnight. The positive colonies were picked and cultured in 5 mL amp-LB medium at 37°C with 250 rpm shaking overnight. Then the amplified plasmid was extracted and identified by *Nde*I and *BamH*I restriction endonuclease enzyme digestion and electrophoresis. The positive plasmid was digested by

*Nde*I and *BamH*I at 37°C overnight. The $GalE_{eco}$ gene fragment was purified by gel extraction and then ligated into pET19b vector to obtain pET19b-$GalE_{eco}$ plasmid. The ligation product was transformed into Jm109 competent cells and cultured on Amp-LB-agar plate at 37°C overnight. The positive colonies were picked and cultured in 5 mL Amp-LB medium at 37°C with 250 rpm shaking overnight. Then, the amplified plasmid was extracted and identified by *Nde*I and *BamH*I restriction endonuclease enzyme digestion and electrophoresis, as well as sequencing. The positive plasmid was then transformed into BL21 Tuner competent cells and cultured on Amp-LB-agar plate at 37°C overnight. The positive colonies were picked and cultured in 5 mL Amp-LB medium at 37°C with 250 rpm shaking overnight, and then stored in 16% glycerol at -80°C.

```
1       ATGAGAGTTC TGGTTACCGG TGGTAGCGGT TACATTGGAA GTCATACCTG
        TGTGCAATTA CTGCAAAACG GTCATGATGT CATCATTCTT GATAACCTCT
101     GTAACAGTAA GCGCAGCGTA CTGCCTGTTA TCGAGCGTTT AGGCGGCAAA
        CATCCAACGT TTGTTGAAGG CGATATTCGT AACGAAGCGT TGATGACCGA
201     GATCCTGCAC GATCACGCTA TCGACACCGT GATCCACTTC GCCGGGCTGA
        AAGCCGTGGG CGAATCGGTA CAAAAACCGC TGGAATATTA CGACAACAAT
301     GTCAACGGCA CTCTGCGCCT GATTAGCGCC ATGCGCGCCG CTAACGTCAA
        AAACTTTATT TTTAGCTCCT CCGCCACCGT TTATGGCGAT CAGCCCAAAA
401     TTCCATACGT TGAAAGCTTC CCGACCGGCA CACCGCAAAG CCCTTACGGC
        AAAAGCAAGC TGATGGTGGA ACAGATCCTC ACCGATCTGC AAAAAGCCCA
501     GCCGGACTGG AGCATTGCCC TGCTGCGCTA CTTCAACCCG GTTGGCGCGC
        ATCCGTCGGG CGATATGGGC GAAGATCCGC AAGGCATTCC GAATAACCTG
601     ATGCCATACA TCGCCCAGGT TGCTGTAGGC CGTCGCGACT CGCTGGCGAT
        TTTTGGTAAC GATTATCCGA CCGAAGATGG TACTGGCGTA CGCGATTACA
701     TCCACGTAAT GGATCTGGCG GACGGTCACG TCGTGGCGAT GGAAAAACTG
        GCGAACAAGC CAGGCGTACA CATCTACAAC CTCGGCGCTG GCGTAGGCAA
801     CAGCGTGCTG GACGTGGTTA ATGCCTTCAG CAAAGCCTGC GGCAAACCGG
        TTAATTATCA TTTTGCACCG CGTCGCGAGG GCGACCTTCC GGCCTACTGG
901     GCGGACGCCA GCAAAGCCGA CCGTGAACTG AACTGGCGCG TAACGCGCAC
        ACTCGATGAA ATGGCGCAGG ACACCTGGCA CTGGCAGTCA CGCCATCCAC
1001    AGGGATATCC CGATTAA
```

Sequence 13.2. Gene sequence of $GalE_{eco}$

#### 13.4.2.2. Expression and purification of UDP-galactose 4-epimerase from *Escherichia coli*

10 $\mu$l BL21T-pET19b-$GalE_{eco}$ was transferred into 250 mL LB medium and cultured at 37°C with 250 rpm shaking overnight. Then, the 250 mL culture was transferred into 5 L TB medium and cultured in a 10 L fermenter at 37°C with 500 rpm stir until OD reached 0.5. 0.1 mM IPTG was added for the expression induction at 28°C for at least 6 hours. After that, the cells were harvested by centrifugation at 4,000 rpm for 20 min and stored at -80°C.

For the purification, 20 g frozen cells were lysed in 200 mL resuspension buffer which contained 0.5 mg/mL lysozyme (Roth, Germany) and 0.5 M NaCl in 20 mM Tris (pH 7.5). The resuspension was treated by ultrasonic for 10 min in ice to obtain crude extract of the cells. After 12,000 rpm centrifugation for 30 min at 4°C, the supernatant (added 25 mM imidazole additionally) was loaded on 5 ml Ni-affinity columns (GE, USA) to purify the recombinant protein. After sample loading, the Ni-affinity columns were washed by 50 mM imidazole wash buffer. Then, the target protein was washed out by 200 mM imidazole eluate buffer. The buffer of the eluate was replaced by 2 mM Tris (pH 7.5) using ultrafiltration. The purity of the recombinant $GalE_{eco}$ was identified by SDS-PAGE electrophoresis and stored in 50% glycerol at -20°C.

### 13.4.3. Chemoenzymatic synthesis of LacNAc and LacNAc-D-T-P-Acr

#### 13.4.3.1. Enzymic synthesis of LacNAc

100 mg GlcNAc (1 eq.) and 384 mg UDP-Glc (1.5 eq.) were added into 50 mL Tris buffer (pH8.0) containing 10 mM $Mn^{2+}$. 1 mL supernatant of cell lysate of BL21T-pET-$GalT_{hpy}$ and 5 mg $GalE_{eco}$ were added to start the reaction. The pH was being adjusted for keeping at 8.0 by pH-stat during the reaction for overnight. When the reaction finished, 300 mL -20°C methanol was added to stop the reaction and precipitate the most of proteins in the reaction system. The precipitate was removed by filtration. The filtrate was concentrated into 10 mL and loaded on P-2 Biogel (Bio-Rad, Germany) column (3 cm × 100 cm). The products were further washed out by $ddH_2O$ and collected separately, and lyophilized for NMR

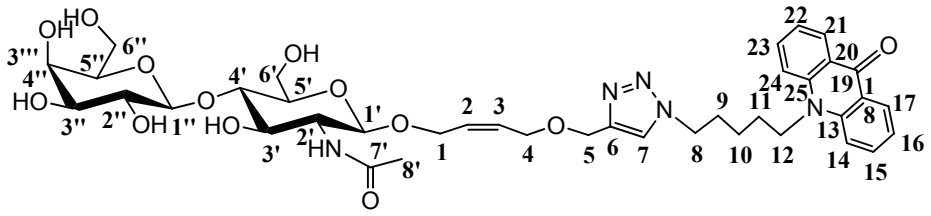
determination. Product: 152 mg white powder. Yield: 88%.

### 13.4.3.2. Chemoenzymatic synthesis of LacNAc-D-T-P-Acr

**10-[5-(4-(((Z)-4-*O*-(β-D-galactopyranosyl-(1,4)-2-acetamido-2-deoxy-β-D-glucopyranosyloxy)-but-2-enyloxy)methyl)-1*H*-1,2,3-triazol-1-yl)pentyl]-10*H*-acridin-9-on**

**(LacNAc-D-T-P-Acr)**

100 mg GlcNAc-D-T-P-Acr (1 eq.) and 133.6 mg UDP-Glc (1.5 eq) were added into 50 mL Gly-Gly buffer (pH8.0) containing 10 mM $Mn^{2+}$. 5 mL supernatant of cell lysate of BL21T-pET-$GalT_{hpy}$ and 5 mg $GalE_{eco}$ were added to start the reaction. The pH was being adjusted for keeping at 8.0 by pH-stat during the reaction for overnight. When the reaction finished, 300 mL -20°C methanol was added to stop the reaction and precipitate the most of proteins in the reaction system. The precipitate was removed by filtration. The filtrate was concentrated into 10 mL and loaded on Biogel P-2 (Bio-Rad, Germany) column (3 cm × 100 cm). The products were further washed out by $ddH_2O$ and collected separately, and lyophilized for NMR determination. Product: 115 mg yellow powder. Yield: 92%.



Chemical Formula: $C_{39}H_{51}N_5O_{13}$
Molecular Weight: 797.85

$^1$H NMR (500 MHz, Deuterium Oxide) δ=7.74 – 7.62 (m, 2H, 17-, 21-H), 7.57 (s, 1H, 7-H), 7.09 (m, 2H, 15-, 23-H), 6.72 (d, *J* = 8.9 Hz, 2H, 14-, 24-H), 6.64 (t, *J* = 7.3 Hz, 2H, 16-, 22-H), 5.52-5.38 (m, 2H, 2-, 3-H), 4.57 (s, 2H, 5-H), 4.38 – 4.23 (m, 4H, 8-, 12-H), 4.10 – 3.89 (m, 4H, 1a-, 1”-, 1’-, 1b-H), 3.88 – 3.69 (m, 3H, 4-, 6’a-H), 3.69 – 3.44 (m, 9H, 6’b-, 4”-, 6”a-, 6”b-, 3”-, 2”-, 4’-, 5”-, 3’-H), 3.44 – 3.35 (m, 2H, 5’-, 2’-H), 1.81 (s, 3H, 8’-H), 1.45-1.25 (m, 2H, 11-H), 1.10-0.65 (m, 4H, 11-, 10-H) ppm.

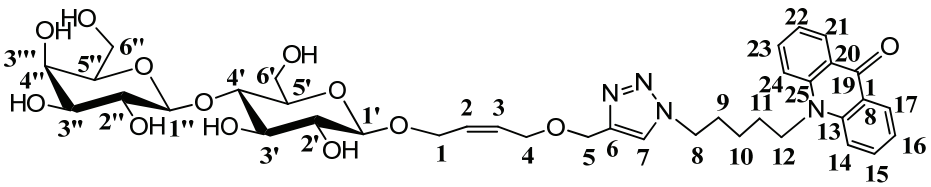
$^{13}$C NMR (126 MHz, Deuterium Oxide) δ=177.79 (C-7'), 174.11 (C-19), 143.90 (C-6), 140.58 (C-15), 134.36 (C-23), 129.59 (C-2, -3), 128.78 (C-17, -21), 126.26 (C-13, -25), 124.47 (C-7), 121.42 (C-16), 120.51 (C-22), 114.88 (C-14, -24), 103.01 (C-1"), 99.90 (C-1'), 78.69 (C-4'), 75.42 (C-3"), 74.87 (C-5"), 73.64 (C-5'), 72.62 (C-3'), 71.03 (C-2'), 69.53 (C-2"), 68.61 (C-4"), 65.33 (C-4), 64.55 (C-1), 62.40 (C-5), 61.05 (C-6"), 60.20 (C-6'), 49.93 (C-8), 45.23 (C-12), 28.91 (C-9), 25.78 (C-8'), 22.70 (C-11), 22.33 (C-10) ppm.

### 13.4.3.3. Chemoenzymatic synthesis of Lac-D-T-P-Acr

**10-[5-(4-(((Z)-4-*O*-(β-D-lactosyloxy)-but-2-enyloxy)methyl)-1*H*-1,2,3-triazol-1-yl)pentyl]-10*H*-acridin-9-on**

**(Lac-D-T-P-Acr)**

For the characterization $GalT_{hpy}$ activity, as well as the structure of the product catalyzed by recombinant $GalT_{hpy}$, 50 mg Glc-D-T-P-Acr (1.5 eq.) and 71 mg UDP-Glc (1.5 eq) were added into 50 mL Gly-Gly buffer (pH8.0) containing 10 mM $Mn^{2+}$. 1 mL supernatant of cell lysate of BL21T-pET-$GalT_{hpy}$ and 5 mg $GalE_{eco}$ were added to start the reaction. The pH was being adjusted for keeping at 8.0 by pH-stat during the reaction for overnight. When the reaction finished, 300 mL -20°C methanol was added to stop the reaction and precipitate the most of proteins in the reaction system. The precipitate was removed by filtration. The filtrate was concentrated into 10 mL and loaded on Biogel P-2 (Bio-Rad, Germany) column (3 cm × 100 cm). The products were further washed out by $ddH_2O$ and collected separately, and lyophilized for NMR determination. Product: 54 mg yellow powder. Yield: 85%.



Chemical Formula: $C_{37}H_{48}N_4O_{13}$
Molecular Weight: 756.80

$^1$H-NMR (500 MHz, $D_2O$): d = 8.45 (dd, *J* = 1.6, 8.0, 2H, 17-, 21-H.), 8.00 (s, 1H, 7-H), 7.85 – 7.82 (m, 2H, 15-, 23-H), 7.72 (d, *J* = 8.7 Hz, 2H, 14-, 24-H), 7.34 (t, *J* = 7.3 Hz, 2H, 16-,

22-H), 5.80 – 5.70 (m, 2H, 2-, 3-H), 4.61 (s, 2H, 5-H), 4.48 – 4.45 (m, 4H, 8- 12-H), 4.40 – 4.38 (m, 1H, 1a-H), 4.37 (d, $J$ = 7.7, 1H, 1”-H), 4.34 (d, $J$ = 7.8, 1H, 1’-H), 4.31 – 4.29 (m, 1H, 1b-H), 4.16 (d, $J$ = 2.8, 2H, 4-H), 3.90 (dd, 1H, 6’a-H), 3.85 – 3.78 (m, 3H, 6’b-, 4”-, 6”a-H), 3.72 – 3.69 (m, 1H, 6”b-H), 3.60 – 3.49 (m, 5H, 3”-, 2”-, 4’-, 5”-, 3’-H), 3.40 – 3.39 (m, 1H, 5’-H), 3.29 – 3.25 (m, 1H, 2’-H), 2.06 – 2.03 (m, 2H, 9-H), 1.94 – 1.91 (m, 2H, 11-H), 1.57 – 1.54 (m, 2H, 10-H) ppm.

$^{13}$C-NMR (126 MHz, $D_2O$): d = 180.16 (C-19), 143.56 (C-6), 136.20 (C-15, -23), 130.76 (C-2, -3), 128.64 (C-17, -21), 123.44 (C-13, -25), 123.11 (C-16, -22),117.14 (C-14, -24), 105.53 (C-1”), 103.49 (C-1’), 81.16 (C-4’), 77.50 (C-3”), 76.93 (C-5”), 76.89 (C-5’), 75.25 (C- 3’), 75.10 (C-2’), 72.97 (C-2”), 70.71 (C-4”), 67.29 (C-4), 65.99 (C-1), 64.47 (C-5), 62.91 (C-6”), 62.41 (C-6’), 51.51 (C-8), 47.19 (C-12), 31.29 (C-9), 28.09 (C-11), 24.87 (C-10) ppm.

## 13.5. Cloning, Expression and characterization of α-2,3- and 2,6-Sialyltransferases from *Photobactirium* sp.

### 13.5.1. Cloning, expression and characterization of α-2,3-sialyltransferase from *Photobacterium phosphoreum*

#### 13.5.1.1. Cloning of α-2,3-sialyltransferase from *Photobacterium phosphoreum*

The DNA sequence of 2,3SiaT$_{pph}$ gene from *Photobacterium phosphoreum* was optimized for expression in *E. coli* and synthesized by Mr. Gene Company (Germany). *Nde*I and *Bam*HI restriction enzyme sites were added at 5'- and 3'- terminus for further cloning work, respectively. The plasmid pMA-PPST from Mr. Gene Company containing synthesized 2,3SiaT$_{pph}$ gene was transformed into *E. coli* Jm109 competent cells for plasmid amplification. The recombinant cells were cultured at 37°C with 250 rpm shaking overnight. The plasmid was extracted and digested by *Nde*I and *Bam*HI, and purified by gel extraction. Then, the 2,3SiaT$_{pph}$ gene fragment was ligated into pET19b vector to construct the expression plasmid pET19b-SiaT$_{pph}$, which was then transformed into *E. coli* Jm109 competent cells for plasmid amplification and identification. The positive colonies were picked and cultured in 5 mL Amp-LB medium at 37°C with 250 rpm shaking overnight. The extracted plasmids were identified by *Nde*I and *Bam*HI restriction endonuclease enzyme digestion and electrophoresis, as well as sequencing. Then the identified final plasmid was transformed into *E. coli* BL21 Tuner competent cells. The positive colony (BL21T-pET19b-SiaT$_{pph}$) was picked and cultured in 5 mL Amp-LB medium at 37°C with 250 rpm shaking overnight and stored in 16% glycerol at -80°C.

```
1       CATATGGACT CGAAACACAA CAACTCTGAC GGCAACATCA CCAAAAACAA
        AACGATTGAG GTCTATGTGG ACCGTGCTAC ACTGCCTACA ATTCAGCAGA
101     TGACCCAGAT TATCAACGAG AACAGTAACA ACAAAAAACT GATTTCCTGG
        AGCCGCTATC CTATTAACGA CGAAACCCTG CTGGAAAGCA TTAATGGCTC
201     ATTCTTCAAA AATCGCCCGG AGCTGATTAA AAGCCTGGAC TCCATGATTC
        TGACCAACGA GATCAAAAAA GTCATTATTA ACGGCAATAC GCTGTGGGCT
301     GTCGACGTGG TAAACATTAT CAAAAGCATC GAGGCACTGG GCAAAAAAAC
        CGAGATCGAA CTGAACTTCT ATGATGATGG TAGCGCCGAG TATGTTCGCC
```

```
401     TGTATGATTT CTCTCGCCTG CCGGAAAGTG AACAGGAGTA TAAAATCTCC
        CTGTCGAAAG ACAACATCCA GAGCAGCATC AACGGTACTC AGCCGTTCGA
501     TAACAGCATT GAAAACATCT ATGGCTTCAG CCAGCTGTAT CCGACAACCT
        ATCACATGCT GCGTGCCGAC ATCTTTGAAA CCAATCTGCC GCTGACTAGC
601     CTGAAACGTG TCATTTCTAA CAACATCAAA CAAATGAAAT GGGACTATTT
        CACGACCTTT AACAGTCAAC AAAAAAACAA ATTCTATAAC TTCACCGGGT
701     TTAATCCGGA GAAAATCAAA GAACAGTATA AAGCCAGCCC TCACGAAAAC
        TTCATCTTCA TTGGCACCAA TTCTGGTACT GCCACTGCCG AGCAACAAAT
801     CGATATCCTG ACCGAGGCCA AAAAACCTGA TTCCCCGATC ATCACGAATA
        GCATCCAAGG ACTGGACCTG TTTTTCAAAG GTCATCCGTC AGCGACGTAT
901     AACCAGCAGA TCATCGATGC CCATAACATG ATCGAGATTT ATAACAAAAT
        CCCGTTCGAA GCCCTGATTA TGACAGACGC TCTGCCAGAT GCCGTGGGTG
1001    GTATGGGTTC TAGCGTTTTC TTCTCACTGC CAAACACCGT GGAGAACAAA
        TTCATCTTCT ATAAAAGCGA CACCGACATT GAGAATAACG CCCTGATCCA
1101    GGTCATGATT GAGCTGAACA TTGTGAACCG TAACGATGTG AAACTGATCA
        GCGACCTGCA GTAAGGATCC
```

Sequence 13.3. Artificial gene sequence of $SiaT_{pph}$

### 13.5.1.2. Expression and purification of α-2,3-sialyltransferase from *Photobacterium phosphoreum*

10 $\mu$l BL21T-pET19b-$SiaT_{pph}$ was transferred into 250 mL LB medium and cultured at 37°C with 250 rpm shaking overnight. Then the 250 mL culture was transferred into 5 L TB medium and cultured in a 10 L fermenter at 37°C with 500 rpm stir until $OD_{600}$ reached 0.5. 0.1 mM IPTG was added for the expression induction at 28°C for at least 6 hours. After that, the cells were harvested by centrifugation at 4,000 rpm for 20 min and stored at -80°C.

For the purification, 20 g frozen cells were lysed in 200 mL resuspension buffer which contained 0.5 mg/mL lysozyme (Roth, Germany) and 0.5 M NaCl in 20 mM Tris (pH 7.5). The resuspension was treated by ultrasonic for 10 min in ice to obtain crude extract of the cells. After 12,000 rpm centrifugation for 30 min at 4°C, the supernatant (added 25 mM imidazole additionally) was loaded on 5 ml Ni-affinity columns (GE, USA) to purify the recombinant protein. After sample loading, the Ni-affinity columns were washed by 50 mM imidazole wash buffer. Then, the target protein was washed out by 200 mM imidazole eluate buffer. The buffer of the eluate was replaced by 2 mM Tris (pH 7.5) using ultrafiltration. The

purity of the recombinant $SiaT_{pph}$ was identified by SDS-PAGE electrophoresis and stored in 50% glycerol at -20°C.

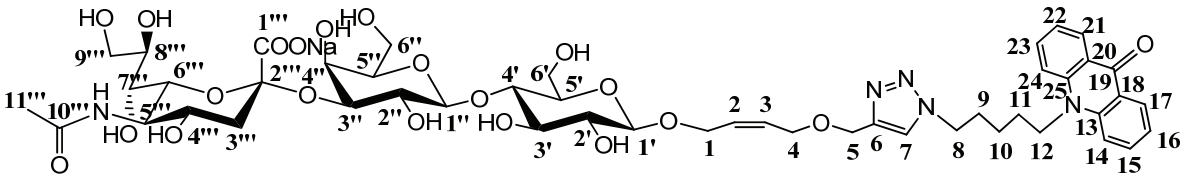
### 13.5.1.3. Characterization of α-2,3-sialyltransferase from *Photobacterium phosphoreum*

The rapid identification of recombinant $SiaT_{pph}$ used a 100 $\mu$l reaction which contained 10 mM Lac-D-T-P-Acr as acceptor, 10 mM CMP-Neu5Ac as donor, 20 mM Tris-HCl (pH8.6) with 0.5 M NaCl and 0.1% Triton X100 as buffer. 50 $\mu$l supernatant of cell lysate of BL21T-pET-$SiaT_{pph}$ was added for the reaction. After 30 min, the Neu5Ac-2,3-Lac-D-T-P-Acr product was identified on TLC plate with mobile phase n-BuOH:aceton:AcOH:$H_2O$=40:40:7:13.

**10-[5-(4-(((Z)-4-*O*-((*N*-acetyl-α-D-neuraminic acid)-(2→3)-β-D-lactosyloxy)-but-2-enyloxy)methyl)-1*H*-1,2,3-triazol-1-yl)pentyl]-10*H*-acridin-9-on**

**(Neu5Ac-2,3- Lac-D-T-P-Acr)**

For the structure identification of the product catalyzed by recombinant $SiaT_{pph}$, 20 mg Lac-D-T-P-Acr (1 eq.) and 25 mg CMP-Neu5Ac (1.5 eq.) were added into 50 mL Tris-HCl (pH8.6) buffer with 0.5 M NaCl and 0.1% Triton × 100. 10 mg recombinant $SiaT_{pph}$ was added to start the reaction. The pH was being adjusted for keeping at 8.6 by pH-stat during the reaction for 2-3 h. When the reaction finished, 300 mL -20°C methanol was added to stop the reaction and precipitate the most of proteins in the reaction system. The precipitate was removed by filtration. The filtrate was concentrated to 10 mL and loaded on Biogel P-2 (Bio-Rad, Germany) column (3 cm × 100 cm). The products were further washed out by $ddH_2O$ and collected and lyophilized for NMR determination.



Chemical Formula: $C_{48}H_{64}N_5NaO_{21}$
Molecular Weight: 1070.03

$^{1}$H NMR (500 MHz, Deuterium Oxide) δ=7.78 (dd, *J* = 8.0, 1.6 Hz, 2H, 17-, 21-H), 7.73 (s, 1H, 7-H), 7.30 – 7.26 (m, 2H, 15-, 23-H), 6.93 – 6.77 (m, 4H, 24-, 14-, 16-, 22-H), 5.70 – 5.48 (m, 2H, 2-, 3-H), 4.45 (s, 2H, 5-H), 4.26 (d, *J* = 8.0 Hz, 1H, 1”-H), 4.23 – 4.06 (m, 4H, 1’-, 1a-, 8-H), 4.05 (dd, *J* = 9.9, 3.1 Hz, 1H, 1b-H), 3.95 – 3.90 (m, 3H, 3”-, 2-, 4-H), 3.87 – 3.75 (m, 5H, 4”-, 6”’-, 9”’a-, 6’-, 5”’-H), 3.73 – 3.46 (m, 13H, 6’b-, 6”-, 8”’-, 7”’-, 9”’b-, 3’-, 4”’-, 4’-, 2”-, 8-, 5”-H), 3.37 – 3.28 (m, 1H, 5’-H), 3.20 (dd, *J* = 9.2, 8.0 Hz, 1H, 2’-H), 2.72 (dd, *J* = 12.4, 4.7 Hz, 1H, 3”’a-H), 1.98 (s, 3H, 11”’-H), 1.75 (t, *J* = 12.1 Hz, 1H, 3”’b-H), 1.56 – 1.50 (m, 2H, 9-H), 1.15 – 0.87 (m, 4H, 11-, 10-H) ppm.

$^{13}$C NMR (126 MHz, Deuterium Oxide) δ=178.13 (C-19), 174.90 (C-1”’), 173.47 (C-10”’), 143.79 (C-6), 140.59 (C-13, -25), 134.57 (C-15, -23), 129.23 (C-2), 129.00 (C-3), 126.18 (C-17, -21), 124.65 (C-7), 121.66 (C-16, -22), 120.33 (C-18, -20), 115.02 (C-14, -24), 103.28 (C-1”), 100.92 (C-1’), 100.31 (C-2”’), 79.59 (C-4’), 74.63 (C-3”), 73.69 (C-8”’), 72.63 (C-5’), 72.55 (C-5”), 72.42 (C-3’), 71.80 (C-2’), 70.79 (C-6”’), 68.52 (C-2”), 68.40 (C-7”’), 68.36 (C-4”’), 65.05 (C-4), 64.63 (C-1), 63.50 (C-9”’), 62.67 (C-5), 62.27 (C-6”), 60.22 (C-6’), 51.83 (C-5”’), 50.05 (C-8), 45.19 (C-12), 40.14 (C-3”’), 28.77 (C-9), 25.73 (C-11), 22.60 (C-10), 22.09 (C-11”’) ppm.

## 13.5.2. Cloning, expression and characterization of α-2,6-sialyltransferase from *Photobacterium leiognathi* JT-SHIZ-145

### 13.5.2.1. Cloning of α-2,6-sialyltransferase from *Photobacterium leiognathi* JT-SHIZ-145

The DNA sequence of 2,6SiaT$_{ple}$ gene from *Photobacterium leiognathi* JT-SHIZ-145 was optimized for expression in *E. coli* and synthesized by the Company Mr. Gene (Germany). *Nde*I and *Bam*HI restriction enzyme sites were added at 5’- and 3’- terminuses for further cloning work, respectively. The plasmid pMA-PLST from Mr. Gene Company containing synthesized 2,6SiaT$_{ple}$ gene was transformed into *E. coli* Jm109 competent cells for plasmid amplification. The recombinant cells were cultured at 37°C with 250 rpm shaking overnight. The plasmid was extracted and digested by *Nde*I and *Bam*HI, and purified by gel extraction. Then, the 2,6SiaT$_{ple}$ gene fragment was ligated into pET19b vector to construct the expression plasmid pET19b-SiaT$_{ple}$, which was then transformed into *E. coli* Jm109 competent cells for plasmid amplification and identification. The positive colonies were picked and cultured in 5

mL Amp-LB medium at 37°C with 250 rpm shaking overnight. The extracted plasmids were identified by *Nde*I and *Bam*HI restriction endonuclease enzyme digestion and electrophoresis, as well as sequencing. Then the identified final plasmid was transformed into *E. coli* BL21 Tuner competent cells. The positive colony (BL21T-pET19b-SiaT$_{ple}$) was picked and cultured in 5 mL Amp-LB medium at 37°C with 250 rpm shaking overnight and stored in 16% glycerol at -80°C.

```
1       GGTACCCATA TGTGTAACGA CAACCAGAAT ACCGTAGACG TGGTGGTGAG
        TACAGTGAAC GATAATGTGA TCGAGAATAA CACCTATCAG GTGAAACCGA
101     TCGATACTCC GACAACGTTT GACTCCTATT CCTGGATTCA GACCTGTGGC
        ACACCAATCC TGAAAGATGA CGAAAAATAT AGCCTGTCCT TCGACTTTGT
201     TGCTCCGGAA CTGGACCAAG ACGAGAAATT TTGCTTCGAG TTTACCGGTG
        ATGTTGATGG AAAACGCTAT GTGACCCAGA CTAATCTGAC CGTGGTAGCA
301     CCGACACTGG AAGTTTATGT GGATCATGCC TCTCTGCCTT CCCTGCAGCA
        GCTGATGAAA ATCATCCAAC AGAAAAACGA ATATTCCCAG AACGAACGCT
401     TTATCAGTTG GGGCCGTATT CGCCTGACGG AGGACAATGC TGAGAAACTG
        AACGCCCACA TTTATCCTCT GGCGGGCAAT AATACAAGCC AAGAGCTGGT
501     CGATGCCGTG ATTGATTATG CCGACTCCAA AAACCGCCTG AACCTGGAGC
        TGAACACTAA TACTGGTCAT TCCTTCCGTA ATATCGCTCC GATCCTGCGT
601     GCAACCTCTT CTAAAAACAA TATCCTGATC AGCAATATCA ACCTGTATGA
        CGATGGTAGC GCCGAATATG TTAGCCTGTA TAACTGGAAA GACACGGACA
701     ATAAAAGCCA AAAACTGAGC GACAGCTTCC TGGTACTGAA AGACTATCTG
        AACGGGATTA GCAGCGAAAA ACCAAATGGG ATCTATAGCA TCTATAACTG
801     GCACCAGCTG TATCATTCGT CCTATTATTT TCTGCGCAAA GATTATCTGA
        CCGTGGAAAC GAAACTGCAC GATCTGCGTG AATATCTGGG CGGATCGCTG
901     AAACAAATGT CATGGGACAC CTTTTCACAA CTGTCCAAAG GCGACAAAGA
        ACTGTTCCTG AACATCGTGG GCTTTGATCA AGAAAAACTG CAGCAAGAAT
1001    ATCAGCAGAG CGAACTGCCG AATTTTGTGT TTACTGGGAC CACAACATGG
        GCAGGTGGCG AGACTAAAGA GTATTATGCC CAGCAGCAGG TGAATGTCGT
1101    GAATAACGCC ATTAACGAAA CGAGCCCTTA TTATCTGGGC CGTGAGCACG
        ACCTGTTCTT CAAAGGTCAT CCTCGTGGGG GTATCATCAA CGACATCATT
1201    CTGGGCAGCT TCAATAACAT GATCGATATC CCGGCCAAAG TGAGTTTTGA
        GGTTCTGATG ATGACCGGAA TGCTGCCTGA TACTGTGGGT GGTATTGCCT
1301    CTTCGCTGTA TTTCTCCATT CCAGCCGAGA AAGTCAGCTT CATCGTGTTT
        ACCAGCAGTG ACACCATTAC AGACCGTGAA GATGCTCTGA AATCACCACT
```

```
1401    GGTTCAAGTC ATGATGACAC TGGGTATCGT GAAAGAGAAA GACGTCCTGT
        TTTGGTGCTG AGGATCCAAG CTT
```

Sequence 13.4. Artificial gene sequence of $SiaT_{ple}$

### 13.5.2.2. Expression and purification of α-2,6-sialyltransferase from *Photobacterium leiognathi* JT-SHIZ-145

10 μl BL21T-pET19b-$SiaT_{ple}$ was transferred into 250 mL LB medium and cultured at 37°C with 250 rpm shaking overnight. Then the 250 mL culture was transferred into 5 L TB medium and cultured in a 10 L fermenter at 37°C with 500 rpm stir until OD reached 0.5. 0.1 mM IPTG was added for the expression induction at 28°C for at least 6 hours. After that, the cells were harvested by centrifugation at 4,000 rpm for 20 min and stored at -80°C.

For the purification, 20 g frozen cells were lysed in 200 mL resuspension buffer which contained 0.5 mg/mL lysozyme (Roth, Germany) and 0.5 M NaCl in 20 mM Tris (pH 7.5). The resuspension was treated by ultrasonic for 10 min in ice to obtain crude extract of the cells. After 12,000 rpm centrifugation for 30 min at 4°C, the supernatant (added 25 mM imidazole additionally) was loaded on 5 ml Ni-affinity columns (GE, USA) to purify the recombinant protein. After sample loading, the Ni-affinity columns were washed by 50 mM imidazole wash buffer. Then, the target protein was washed out by 200 mM imidazole eluate buffer. The buffer of the eluate was replaced by 2 mM Tris (pH 7.5) using ultrafiltration. The purity of the recombinant 2,6$SiaT_{ple}$ was identified by SDS-PAGE electrophoresis and stored in 50% glycerol at -20°C.

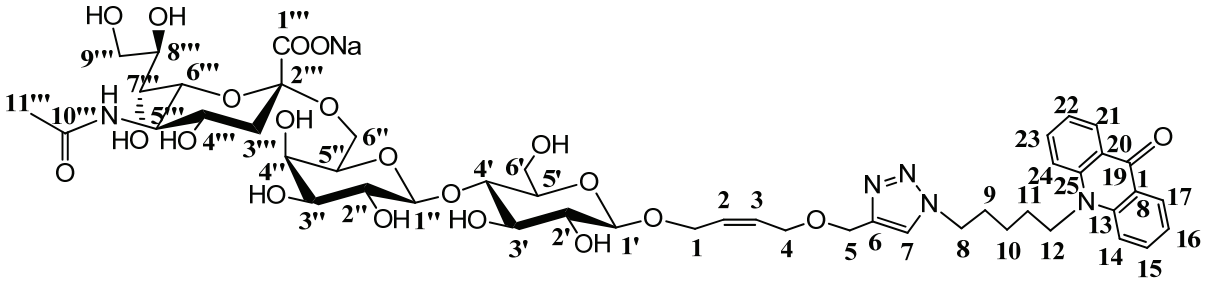
### 13.5.2.3. Characterization of α-2,6-sialyltransferase from *Photobacterium leiognathi* JT-SHIZ-145

The rapid identification of recombinant 2,6$SiaT_{ple}$ used a 100 μl reaction which contained 10 mM Lac-D-T-P-Acr as acceptor, 10 mM CMP-Neu5Ac as donor, 20 mM Tris-HCl (pH8.6) with 0.5 M NaCl and 0.1% Triton X100 as buffer. 50 μl supernatant of cell lysate of BL21T-pET-$SiaT_{pph}$ was added for the reaction. After 30 min, the Neu5Ac-2,6-Lac-D-T-P-Acr product was identified on TLC plate with mobile phase n-BuOH:aceton:AcOH:$H_2O$=40:40:7:13.

**10-[5-(4-(((Z)-4-*O*-((*N*-acetyl-α-D-neuraminic acid)-(2→6)-β-D-lactosyloxy)-but-2-enyloxy)methyl)-1*H*-1,2,3-triazol-1-yl)pentyl]-10*H*-acridin-9-on**

**(Neu5Ac-2,6- Lac-D-T-P-Acr)**

For the structure identification of the product catalyzed by recombinant 2,6SiaT$_{ple}$, 20 mg Lac-D-T-P-Acr (1 eq.) and 25 mg CMP-Neu5Ac (1.5 eq.) were added into 50 mL Tris-HCl (pH8.6) buffer with 0.5 M NaCl and 0.1% Triton ×100. 10 mg recombinant SiaT$_{pph}$ was added to start the reaction. The pH was being adjusted for keeping at 8.6 by pH-stat during the reaction for 2-3 h. When the reaction finished, 300 mL -20°C methanol was added to stop the reaction and precipitate the most of proteins in the reaction system. The precipitate was removed by filtration. The filtrate was concentrated to 10 mL and loaded on Biogel P-2 (Bio-Rad, Germany) column (3 cm × 100 cm). The products were further washed out by dd$H_2O$ and collected and lyophilized for NMR determination.



Chemical Formula: $C_{48}H_{64}N_5NaO_{21}$
Molecular Weight: 1070.03

$^1$H NMR (500 MHz, Deuterium Oxide) δ=7.81 (m, 2H, 17- 21H), 7.74 (s, 1H, 7-H), 7.34 (dd, 2H, 15-, 23-H), 7.02 – 6.82 (m, 4H, 14-, 24-, 16-, 22-H), 5.73 – 5.48 (m, 2H, 2-, 3-H), 4.46 (s, 2H, 5-H), 4.30 (d, *J* = 7.8 Hz, 1H, 1”-H), 4.27 (dd, *J* = 8.0, 1.6Hz, 1H, 1’-H), 4.24-4.13 (m, 3H, 8-, 1a-H), 4.13-4.06 (m, 1H, 1b-H), 3.96 – 3.75 (m, 9H, 4-, 9”’a-, 5”’-, 8”’-, 6”’-, 6’a-, 6”a-, 4”-H), 3.75 – 3.62 (m, 3H, 4”’-, 5”-, 6’b-H), 3.62 – 3.43 (m, 9H, 6”b-, 3”-, 9”’-, 12-, 7”’-, 3’-, 4’-, 2”-H), 3.41 – 3.32 (m, 1H, 5’-H), 3.22 (t, *J* = 8.1 Hz, 1H, 2’-H), 2.66 (dd, *J* = 12.4, 4.6Hz, 1H, 3”’a-H), 1.97 (s, 3H, 11”’-H), 1.67 (t, *J* = 12.2 Hz, 1H, 3”’-H), 1.62-1.50 (m, 2H, 9-H), 1.16 – 0.92 (m, 4H, 11-, 10-H) ppm.

$^{13}$C NMR (126 MHz, Deuterium Oxide) δ=178.13 (C-19), 174.90 (C-1'''), 173.47 (C-10'''), 143.79 (C-6), 140.59 (C-13, -25), 134.57 (C-15, -23), 129.23 (C-2), 129.00 (C-3), 126.18 (C-17, -21), 124.65 (C-7), 121.66 (C-16, -21), 120.33 (C-18, -20), 115.02 (C-14, -24), 103.28 (C-1"), 100.92 (C-1'), 100.31 (C-2'''), 79.59 (C-4'),74.67 (C-3"), 74.63 (C-8'''), 73.69 (C-5'), 72.63 (C-5"), 72.55 (C-3'), 72.42 (C-2'), 71.80 (C-6'''), 70.79 (C-2"), 68.52 (C-7'''), 68.40 (C-4'''), 68.36 (C-4"), 65.05 (C-4), 64.63 (C-1), 63.50 (C-6"), 62.67 (C-9'''), 62.27 (C-5), 60.22 (C-6'), 51.83 (C-5'''), 50.05 (C-8), 45.19 (C-12), 40.14 (C-3'''), 28.77 (C-9), 25.73 (C-11), 22.60 (C-10), 22.09 (C-11''') ppm.

### 13.5.3. Development of pH-based assay for sialyltransferases

#### 13.5.3.1. Calibration curve, limit of detection (LOD) and limit of quantification (LOQ)

The standard curve of the assay method used a series of finial concentration of HCl from 0 mM to 1.0 mM in a 96-well plate. The mixture also contained 1 mM CMP-Neu5Ac, 2 $\mu$g $SiaT_{ple}$, 2 $\mu$l 0.1% phenol red and 2 mM Tris-HCl (pH 8.0). Total volume was 200$\mu$l. The relationship between concentration of proton and absorbance was measured at 560 nm by Vmax plate reader (Beckman, USA) to obtain the standard curve $y = ax + b$. LOD was defined as $LOD = 3.3Sb / \bar{a}$, and LOQ was defined as $LOD = 10Sb / \bar{a}$, where $Sb$ is the standard deviation of the control.

#### 13.5.3.2. Reaction system for assay

The assay reaction was taken place on 96-well flat bottom plate. The reaction mixture contained 0-4 $\mu$l SiaT (dependent on the activity of SiaT), 20 mM acceptor, 2 mM Tris-HCl (pH 8.0) and 2 $\mu$l phenol red (0.1%, 2.82 mM). 1 mM CMP-Neu5Ac was added to start the reaction. The total volume in each well was 200 $\mu$l. The absorbance increase was measured at 560 nm by plate reader at 20°C.

#### 13.5.3.3. Dependence of rate on amount of SiaT

The specific activity measurement at varying enzyme concentration used a series of amount of 0-4 $\mu$g $SiaT_{ple}$. 20 mM Lac was used as the acceptor. Absorbance data for the first 60 sec was recorded and converted into concentration of proton according to the calibration curve for the activity calculation.

### 13.5.4. Substrate tolerance screening of sialyltransferase

#### 13.5.4.1. Determination of acceptor specificity

To screening the acceptor tolerance of SiaTs, the activities towards various acceptors were measured, including 20 mM acceptor, 2 mM Tris-HCl (pH 8.0) and 2 $\mu$l phenol red (0.1% , 2.82 mM). 1 mM CMP-Neu5Ac was added to start the reaction. The total volume in each well was 200 $\mu$l. The absorbance increase was measured at 560 nm by plate reader at 20°C.

#### 13.5.4.2. Determination of kinetic constants

The enzymatic kinetics towards both donor CMP-Neu5Ac and acceptor (galactose, methyl-α-D-galactopyranoside, methyl-β-D-galactopyranoside and lactose) were determined at varying concentration of substrates. 1 mM CMP-Neu5Ac was used for the determination of kinetic constants for the acceptors. 20 mM Lac was used for the measurement of CMP-Neu5Ac kinetics. The reaction rate at each concentration was calculated according to the calculation curve and collected for non-linear fit to the Michaelis-Menten equation using Origin 8 software to obtain enzymatic kinetic values.

## 13.6. Directed Evolution of Cytosine-5′-Monophosphate-*N*-Acetylneuraminate Synthetase from *Neisseria meningitidis* towards 5′-modified *N*-Acetylneuraminate Analogues.

### 13.6.1. Building of Phe192 and Phe193 single site mutagenesis libraries

The construction of Phe192 and Phe193 single-site mutagenesis libraries of CSS used QuikChange Lighning Mutagenesis Kit. PCR primers CSSF192f 5'- C ATT AAT GAT ACT GCT TCA CTA ATT GCA AAT AAT TGT NNK TTT ATC GCC CCA ACC AAA CTT TAT -3' (forward primer) and CSSF192b 5'- ATA AAG TTT GGT TGG GGC GAT AAA MNN ACA ATT ATT TGC AAT TAG TGA AGC AGT ATC ATT AAT G -3' (backward primer) and CSSF193f 5'- C ATT AAT GAT ACT GCT TCA CTA ATT GCA AAT AAT TGT TTT NNK ATC GCC CCA ACC AAA CTT TAT -3' (forward primer) and CSSF193b 5'- ATA AAG TTT GGT TGG GGC GAT MNN AAA ACA ATT ATT TGC AAT TAG TGA AGC AGT ATC ATT AAT G -3' (backward primer) were designed for Phe192 and Phe193 sites, respectively. The PCR used CSS gene in vector pET21a(+) (pET21a-CSS) as template.

PCR reaction:

| | |
|---|---|
| dsDNA template: | 2 $\mu$l |
| Primer 1: | 2 $\mu$l |
| Primer 2: | 2 $\mu$l |
| dNTP mix: | 1 $\mu$l |
| 10×Reaction Buffer: | 5 $\mu$l |
| QuikSolution reagent: | 1.5 $\mu$l |
| QuikChange Lightning Enzyme: | 1 $\mu$l |
| dd$H_2O$: | up to 50 $\mu$l |

| | | |
|---|---|---|
| 95°C | 2 min | |
| 95°C | 30 sec | |
| 68°C | 10 sec | |
| 68°C | 3 min 40 sec | 30 cycles |
| 68°C | 10 min | |

The PCR product was digested by 2 $\mu$l *Dpn*I for 1 h at 37°C. Then 2 $\mu$l PCR product was transformed into BL21DE3 competent cells and cultured on ampicillin agar plates overnight at

37°C. 96 colonies were picked into 96-well plates containing 150 μl/well LB-ampicillin medium. After overnight culture at 37°C 900 rpm, 30 μl glycerol was added into each well. The whole plates were sealed with plastic lids and stored at -80°C.

### 13.6.2. Building of Phe192-Phe193 double site mutagenesis library

The construction of Phe192-Phe193 double-site mutagenesis libraries of CSS used QuikChange Lighning Mutagenesis Kit with two mutagenesis strategies. PCR primers CSSF192193Rf 5'- GCA ATT TAC ATT AAT GAT ACT GCT TCA CTA ATT GCA AAT AAT TGT NRT NRT ATC GCC CCA ACC AAA CTT TAT ATT ATG -3' (forward primer) and CSSF192193Rb 5'- CAT AAT ATA AAG TTT GGT TGG GGC GAT AYN AYN ACA ATT ATT TGC AAT TAG TGA AGC AGT ATC ATT AAT GTA AAT TGC -3' (backward primer) and CSSF192193Df 5'- GCA ATT TAC ATT AAT GAT ACT GCT TCA CTA ATT GCA AAT AAT TGT NDT NDT ATC GCC CCA ACC AAA CTT TAT ATT ATG -3' (forward primer) and CSSF192193Db 5'- CAT AAT ATA AAG TTT GGT TGG GGC GAT AHN AHN ACA ATT ATT TGC AAT TAG TGA AGC AGT ATC ATT AAT GTA AAT TGC -3' (backward primer) were designed for Phe192193R and Phe192193D libraries, respectively. The PCR used CSS gene in vector pET21a(+) (pET21a-CSS) as template.

PCR reaction:

| | |
|---|---|
| dsDNA template: | 2 μl |
| Primer 1: | 2 μl |
| Primer 2: | 2 μl |
| dNTP mix: | 1 μl |
| 10×Reaction Buffer: | 5 μl |
| QuikSolution reagent: | 1.5 μl |
| QuikChange Lightning Enzyme: | 1 μl |
| $ddH_2O$: | up to 50 μl |

| | | |
|---|---|---|
| 95°C | 2 min | |
| 95°C | 30 sec | |
| 68°C | 10 sec | |
| 68°C | 4 min | 30 cycles |
| 68°C | 10 min | |

The PCR process is described as follows: 95°C 2 min, (95°C 10 sec, 68°C 3 min 40 sec) 30 cycles, 68°C 5 min. The PCR product was digested by 2 $\mu$l *Dpn*I for 1 h at 37°C. Then 2 $\mu$l PCR product was transformed into BL21DE3 competent cells and cultured on ampicillin agar plates overnight at 37°C. 192 and 432 colonies were picked into 96-well plates containing 150 $\mu$l/well LB-ampicillin medium for Phe192193R and Phe192193D libraries, respectively. After overnight culture at 37°C 900 rpm, 30 $\mu$l glycerol was added into each well. The whole plates were sealed with plastic lids and stored at -80°C.

### 13.6.3. Screening of CSS mutants

2 $\mu$l of each clone in libraries was transferred to individual wells in 96-well round bottom plates containing 150 $\mu$l/well LB growth media comprised 50 $\mu$g/mL ampicillin. The plates were sealed with plastic lids to avoid evaporation, and then incubated at 30°C, 1,000 rpm overnight. After that, 5 $\mu$l was transferred from each well to 96-deepwell plates with plastic lids containing 400 $\mu$l/well of LB-kanamycin growth medium containing 0.1 mM IPTG. The plates were then incubated at 30°C, 1,000 rpm overnight, until the amount of cells in each well reached the maximum. Cells were harvested by centrifugation at 4,000 rpm for 20 min. The culture medium was removed and the cell pellets could be stored at -80°C.

The cell pellets were resuspended in 150 $\mu$l resuspension buffer (1/10 BugBuster solution (Novagen, Germany), 0.5 mg/mL lysozyme (Roth, Germany) and 4U/mL Benzonase endonuclease (Novagen, Germany)) and incubated in a shaker for 30 min at room temperature. Then, the cell lysate was centrifugated at 4,000 rpm for 30 min. 40 $\mu$l of the supernatant was transferred to a new 96-well flat bottom plate. 140 $\mu$l of substrate solution (2 mM Tris (pH8.6), 10 mM $MgCl_2$, 2.82 mM cresol red and 10 mM acceptor was added to each well. The reaction was initiated by the addition of 0.5 mM CTP. The total volume in each well was 200 $\mu$l. The OD increase was read by the plate-reader at 574 nm.

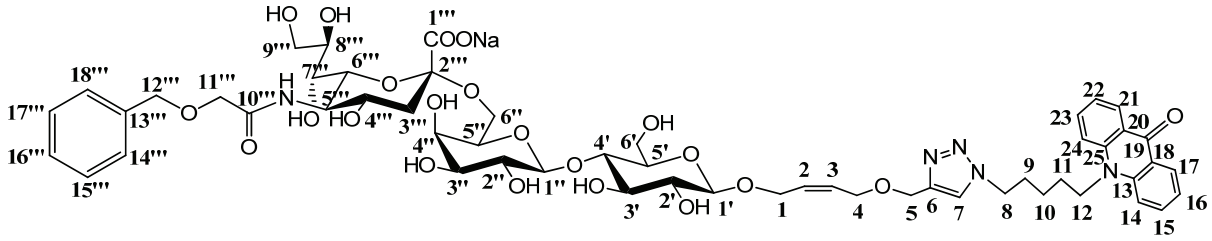
16 (for single site libraries) or 24 (for double site libraries) positive colonies with the highest activity were picked from each plate into a new plate for the second round screening.

## 13.6.4. Verification by synthesis of neo-sialoconjugates

**10-[5-(4-(((Z)-4-*O*-((*N*-benzyloxy-acetyl -α-D-neuraminic acid)-(2→6)-β-D-lactosyloxy)-but-2-enyloxy)methyl)-1*H*-1,2,3-triazol-1-yl)pentyl]-10*H*-acridin-9-on**

**(Neu5AcOBz-2,6-Lac-D-T-P-Acr)**

For the structure identification of the product catalyzed by recombinant 2,6SiaT$_{ple}$, 25.2 mg CTP (2 eq.) and 17.3 mg Neu5AcOBz (1.5 eq.) were added into 50 mL Tris-HCl (pH8.6) buffer with 0.5 M NaCl and 0.1% Triton ×100. 5 mg CSS-F192C/F193Y and 20 μl PPase were added to start the reaction. When the CMP-activation was completed, 20 mg Lac-D-T-P-Acr (1 eq.) and 5 mg recombinant SiaT$_{ple}$ were added to start the reaction. The pH was being adjusted for keeping at 8.6 by pH-stat during the reaction for 2-3 h. When the reaction finished, 300 mL -20°C methanol was added to stop the reaction and precipitate the most of proteins in the reaction system. The precipitate was removed by filtration. The filtrate was concentrated to 10 mL and loaded on Biogel P-2 (Bio-Rad, Germany) column (3 cm × 100 cm). The products were further washed out by dd$H_2O$ and collected and lyophilized for NMR determination. Product: 15.5 mg yellow powder. Yield: 50%.



Chemical Formula: $C_{55}H_{70}N_5NaO_{22}$
Molecular Weight: 1176.15

$^1$H NMR (500 MHz, $D_2O$) δ=7.88 (dd, *J* = 8.0, 1.1 Hz, 2H, 17-, 21-H), 7.74 (s, 1H, 7-H), 7.38 (dd, *J* = 8.3, 6.3 Hz, 2H, 15-, 23-H), 7.29 – 7.18 (m, 5H, 14'''-, 15'''-, 16'''-, 17'''-, 18'''-H), 7.02 – 6.92 (m, 4H, 14-, 24-, 16-, 22-H), 5.75 – 5.57 (m, 2H, 2-, 3-H), 4.47 (d, *J* = 6.5 Hz, 4H, 5-, 12'''-H), 4.34 (d, *J* = 7.8 Hz, 1H, 1''-H), 4.31 (d, *J* = 8.0 Hz, 1H, 1'-H), 4.21 (dt, *J* = 18.7, 6.3 Hz, 3H, 1a-, 8-H), 4.12 (dd, *J* = 12.4, 7.1 Hz, 1H, 1b-H), 4.01 – 3.91 (m, 4H, 11'''a-, 4-, 9'''a-H), 3.91 – 3.78 (m, 6H, ), 3.77 – 3.63 (m, 6H, 6'''-, 8'''-, 6'a-, 6''a-, 5'''-H), 3.62 – 3.44 (m, 12H, 5''-, 6'b-, 4'''-, 3''-, 6''b-, 9'''b-, 4''-, 12-, 3'-, 4'-, 2''-, 7'''-H), 3.44 – 3.38 (m, 1H, 5'-H), 3.25

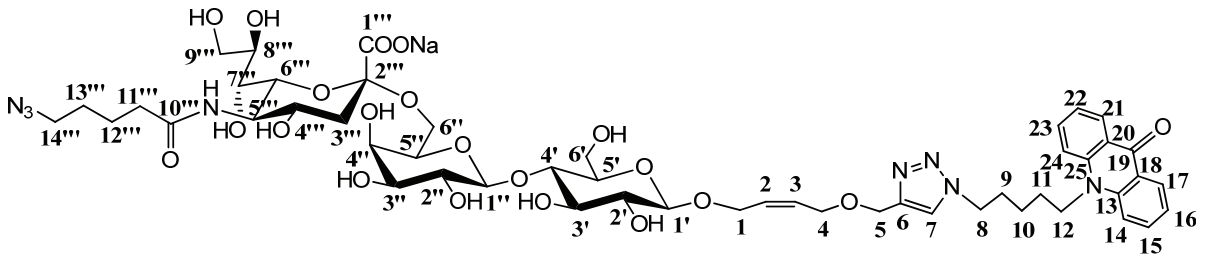
(t, *J* = 8.5 Hz, 1H, 2'-H), 2.68 (dd, *J* = 12.4, 4.6 Hz, 1H, 3'''a-H), 1.71 (t, *J* = 12.2 Hz, 1H, 3'''b-H), 1.63 – 1.52 (m, 2H, 9-H), 1.20 – 1.07 (m, 2H, 11-H), 1.06 – 0.96 (m, 2H, 10-H) ppm.

$^{13}$C NMR (126 MHz, $D_2O$) δ=178.24 (C-19), 173.49 (C-1'''), 173.16 (C-10'''), 143.83 (C-6), 140.68 (C-13, -25), 136.50 (C-13'''), 134.62 (C-15, -23), 129.35 (C-2), 128.94 (C-3), 128.68 (C-14''', -18'''), 128.43 (C-15''', -17''', -16'''), 126.26 (C-17, -21), 124.68 (C-7), 121.71 (C-16, -22), 120.42 (C-18, -20), 115.09 (C-14, -24), 103.20 (C-1''), 101.03 (C-1'), 100.22 (C-2'''), 79.66 (C-4'), 74.68 (C-5'), 74.63 (C-3'), 73.77 (C-5''), 73.22, 72.66 (C-2'), 72.44 (C-4''), 72.27 (C-3''), 71.91 (C-8'''), 70.78 (C-2''), 68.52 (C-7'''), 68.44 (C-11'''), 68.04 (C-4'''), 65.10 (C-4), 64.69 (C-1), 63.61 (C-9'''), 62.73 (C-6''), 62.28 (C-5), 60.28 (C-6'), 51.65 (C-5'''), 50.06 (C-8), 45.21 (C-12), 40.33 (C-3'''), 28.80 (C-9), 25.79 (C-11), 22.63 (C-10) ppm.

**10-[5-(4-(((Z)-4-*O*-((*N*-azidopentanoyl -α-D-neuraminic acid)-(2→3)-β-D-lactosyloxy)-but-2-enyloxy)methyl)-1*H*-1,2,3-triazol-1-yl)pentyl]-10*H*-acridin-9-on**

**( NeuPen$N_3$-2,6-Lac-D-T-P-Acr)**

For the structure identification of the product catalyzed by recombinant 2,6Sia$T_{ple}$, 25.2 mg CTP (2 eq.) and 16.4 mg NeuPen$N_3$ (1.5 eq.) were added into 50 mL Tris-HCl (pH8.6) buffer with 0.5 M NaCl and 0.1% Triton ×100. 5 mg CSS-F192S and 20 μl PPase were added to start the reaction. When the CMP-activation was completed, 20 mg Lac-D-T-P-Acr (1 eq.) and 5 mg recombinant Sia$T_{ple}$ were added to start the reaction. The pH was being adjusted for keeping at 8.6 by pH-stat during the reaction for 2-3 h. When the reaction finished, 300 mL -20°C methanol was added to stop the reaction and precipitate the most of proteins in the reaction system. The precipitate was removed by filtration. The filtrate was concentrated to 10 mL and loaded on Biogel P-2 (Bio-Rad, Germany) column (3 cm × 100 cm). The products were further washed out by dd$H_2O$ and collected and lyophilized for NMR determination. Product: 19.5 mg yellow powder. Yield: 64%.



Chemical Formula: $C_{51}H_{69}N_8NaO_{21}$
Molecular Weight: 1153.12

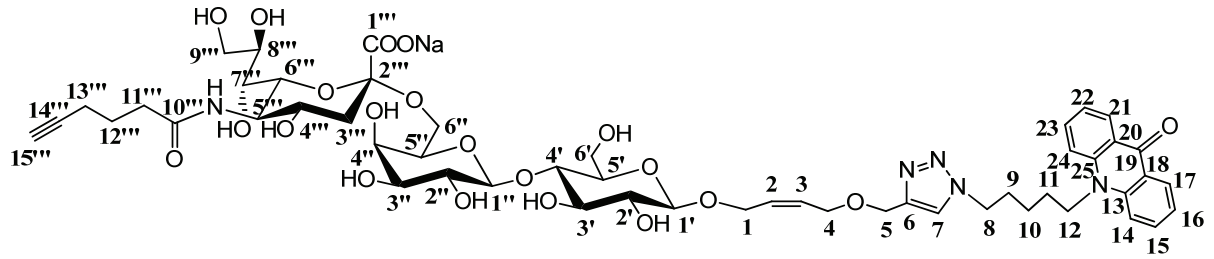
$^{1}$H NMR (500 MHz, $D_2O$) δ=7.84 (t, *J* = 8.9 Hz, 2H, 17-, 21-H), 7.74 (s, 1H, 7-H), 7.34 (t, *J* = 7.3 Hz, 2H, 15-, 23-H), 7.01 – 6.86 (m, 4H, 14-, 24-, 16-, 22-H), 5.74 – 5.56 (m, 2H, 2-, 3-H), 4.46 (d, *J* = 4.2 Hz, 2H, 5-H), 4.33 (t, *J* = 7.7 Hz, 1H, 1”-H), 4.30 (d, *J* = 8.0 Hz, 1H, 1’-H), 4.25 – 4.15 (m, 3H, 1a-, 8-H), 4.11 (dd, *J* = 12.6, 7.0 Hz, 1H, 1b-H), 3.98 – 3.89 (m, 3H, 4-, 9”á-H), 3.90 – 3.77 (m, 5H, 6”’-, 8”’-, 6’a-, 6”a-, 5”’-H), 3.77 – 3.45 (m, 13H, 5”-, 6’b-, 4”-, 4”’-, 3”-, 6”b-, 12-, 9”’b-, 3’-, 7”’-, 4’-, 2”-H), 3.42 – 3.36 (m, 1H, 5’-H), 3.22 (dt, *J* = 10.3, 7.5 Hz, 3H, 2’-, 14”’-H), 2.67 (dd, *J* = 12.3, 4.6 Hz, 1H, 3”’a-H), 2.23 (t, *J* = 7.2 Hz, 2H, 11”’-H), 1.70 (t, *J* = 12.2 Hz, 1H, 3”’b-H), 1.64 – 1.53 (m, 4H, 9-, 12”’-H), 1.52 – 1.45 (m, 2H, 13”’-H), 1.17 – 1.06 (m, 2H, 11-H), 1.05 – 0.97 (m, 2H, 10-H) ppm.

$^{13}$C NMR (126 MHz, $D_2O$) δ=178.12 (C-19), 177.28 (C-1”’), 173.36 (C-10”’), 143.82 (C-6), 140.65 (C-13, -25), 134.57 (C-15, -23), 129.30 (C-2), 128.96 (C-3), 126.25 (C-17, -21), 124.64 (C-7), 121.64 (C-16, -22), 120.41 (C-18, -20), 115.04 (C-14, -24), 103.26 (C-1”), 100.99 (C-1’), 100.19 (C-2”’), 79.65 (C-4’), 74.69 (C-3’), 74.64 (C-5’), 73.71 (C-5”), 72.65 (C-4”), 72.61 (C-3”), 72.44 (C-2’), 71.76 (C-8”’), 70.78 (C-2”), 68.56 (C-7”’), 68.46 (C-6”’), 68.15 (C-4”’), 65.11 (C-4), 64.67 (C-1), 63.52 (C-9”’), 62.72 (C-6”), 62.30 (C-5), 60.27 (C-6’), 51.78 (C-5”’), 50.70 (C-14”’), 50.04 (C-8), 45.21(C-12), 40.27 (C-3”’), 35.30 (C-11”’), 28.82 (C-9), 27.50 (C-13”’), 25.79 (C-11), 22.64 (C-10), 22.56 (C-12”’) ppm.

**10-[5-(4-(((Z)-4-*O*-((*N*-(hex-5-ynoyl)-α-D-neuraminic acid)-(2→6)-β-D-lactosyloxy)-but-2-enyloxy)methyl)-1*H*-1,2,3-triazol-1-yl)pentyl]-10*H*-acridin-9-on**

**(Neu5Hex-2,6-Lac-D-T-P-Acr)**

For the structure identification of the product catalyzed by recombinant 2,6SiaT$_{ple}$, 25.2 mg CTP (2 eq.) and 15.2 mg Neu5Hex (1.5 eq.) were added into 50 mL Tris-HCl (pH8.6) buffer with 0.5 M NaCl and 0.1% Triton ×100. 5 mg CSS-F192S/F193Y and 20 $\mu$l PPase were added to start the reaction. When the CMP-activation was completed, 20 mg Lac-D-T-P-Acr (1 eq.) and 5 mg recombinant SiaT$_{ple}$ were added to start the reaction. The pH was being adjusted for keeping at 8.6 by pH-stat during the reaction for 2-3 h. When the reaction finished, 300 mL -20°C methanol was added to stop the reaction and precipitate the most of proteins in the reaction system. The precipitate was removed by filtration. The filtrate was concentrated to 10 mL and loaded on Biogel P-2 (Bio-Rad, Germany) column (3 cm × 100 cm). The products were further washed out by ddH2O and collected and lyophilized for NMR determination. Product: 20.7 mg yellow powder. Yield: 70%.



Chemical Formula: $C_{52}H_{68}N_5NaO_{21}$
Molecular Weight: 1122.11

$^{1}$H NMR (500 MHz, Deuterium Oxide) δ 7.84 (dd, *J* = 8.1, 1.6 Hz, 2H, 17-, 21-H), 7.75 (s, 1H, 7-H), 7.36 (ddd, *J* = 8.4, 6.7, 1.6 Hz, 2H, 15-, 23-H), 7.01 – 6.88 (m, 4H, 14-, 24-, 16-, 22-H), 5.71 – 5.54 (m, 2H, 5-H), 4.47 (s, 2H, 2-, 3-H), 4.32 (d, *J* = 7.8 Hz, 1H, 1”-H), 4.28 (d, *J* = 8.0 Hz, 1H, 1’-H), 4.24 – 4.16 (m, 3H, 1a-, 8-H), 4.10 (dd, *J* = 12.6, 6.9 Hz, 1H, 1b-H), 3.99 – 3.88 (m, 3H, 4-, 9’”a-H), 3.88 – 3.76 (m, 6H, 6’”-, 8’”-, 6a-, 6”a-, 5’”-H), 3.76 – 3.43 (m, 17H, 5”-, 6’b-, 4’”-, 3”-, 6”b-, 9’”b-, 3’-, 12-, 4’-, 7’”-, 2”-H), 3.38 (ddd, *J* = 9.3, 5.0, 2.3 Hz, 1H, 5’-H), 3.23 (t, *J* = 8.5 Hz, 1H, 2’-H), 2.66 (dd, *J* = 12.4, 4.7 Hz, 1H, 3’”a-H), 2.33 (ddd, *J* = 8.1, 6.0, 2.1 Hz, 2H, 11’”-H), 2.14 (tt, *J* = 6.9, 2.1 Hz, 2H, 13’”-H), 1.75 – 1.64 (m, 3H, 12’”-, 3’”b-H), 1.59 (t, *J* = 7.4 Hz, 2H, 9-H), 1.19 – 0.93 (m, 4H, 11-, 10-H) ppm.

$^{13}$C NMR (126 MHz, Deuterium Oxide) δ 178.19 (C-19), 176.85 (C-1’”), 173.50 (C-10’”), 140.64 (C-13, -25), 134.61 (C-15, -23), 129.28 (C-2), 128.98 (C-3), 126.22 (C-17, -21), 124.68 (C-7), 121.69 (C-16, -22), 120.38 (C-18, -20), 115.07 (C-14, -24), 103.26 (C-1”), 100.96 (C-1’), 100.28 (C-2’”), 84.16 (C-14’”), 79.63 (C-4’), 74.69 (C-5’), 74.63 (C-3’), 73.72 (C-5”), 72.64 (C-2’), 72.55 (C-4”), 72.43 (C-3”), 71.87 (C-8’”), 70.79 (C-2”), 69.98 (C-15’”), 68.57 (C-7’”), 68.53 (C-6’”), 68.26 (C-4’”), 65.06 (C-4), 64.66 (C-1), 63.53 (C-9’”), 62.71 (C-6”), 62.28 (C-5), 60.25 (C-6’), 51.77 (C-5’”), 50.07 (C-8), 45.21 (C-12), 40.29 (C-3’”), 34.61 (C-11’”), 28.79 (C-9), 25.77 (C-11), 24.01 (C-12’”), 22.63 (C-10), 17.02(C-13’”) ppm.

## 13.7. Directed Evolution of 2,6-Sialyltransferase from *Photobacterium leiognathi* JT-SHIZ-145 towards α-Galactosides

### 13.7.1. Building of Trp347 site mutagenesis library

The construction of Trp347 single-site mutagenesis library of 2,6SiaTple used QuikChange Lighning Mutagenesis Kit. PCR primers PLSTW347f 5'- CG AAT TTT GTG TTT ACT GGG ACC ACA ACA NNS GCA GGT GGC GAG ACT -3' (forward primer) and PLSTW347b 5'- AGT CTC GCC ACC TGC SNN TGT TGT GGT CCC AGT AAA CAC AAA ATT CG -3' (backward primer) were designed for Trp347 site. The PCR used artificial 2,6SiaT$_{ple}$ gene, which has been optimized for the expression in *E. coli*, in vector pET19b (pET19-PLST) as template.

PCR reaction:

| | |
|---|---|
| dsDNA template: | 2 μl |
| Primer 1: | 2 μl |
| Primer 2: | 2 μl |
| dNTP mix: | 1 μl |
| 10×Reaction Buffer: | 5 μl |
| QuikSolution reagent: | 1.5 μl |
| QuikChange Lightning Enzyme: | 1 μl |
| dd$H_2O$: | up to 50 μl |

| | | |
|---|---|---|
| 95°C | 2 min | |
| 95°C | 30 sec | |
| 72°C | 10 sec | |
| 68°C | 4 min | 30 cycles |
| 68°C | 10 min | |

The PCR product was digested by 2 μl *Dpn*I for 1 h at 37°C. Then 2 μl PCR product was transformed into BL21DE3 competent cells and cultured on ampicillin agar plates overnight at 37°C. 192 colonies were picked into 96-well plates containing 150 μl/well LB-ampicillin medium. After overnight culture at 37°C 900 rpm, 30 μl glycerol was added into each well. The whole plates were sealed with plastic lids and stored at -80°C.

### 13.7.2.Building of T345 site mutagenesis library

The construction of Trp347Gly and Trp347Ala single-site mutants of 2,6SiaTple used QuikChange Lighning Mutagenesis Kit. PCR primers PLSTW347Gf 5'- G TTT ACT GGG ACC ACA ACA GGG GCA GGT GG -3' (forward primer) and PLSTW347Gb 5'- CC ACC TGC CCC TGT TGT GGT CCC AGT AAA C -3' (backward primer); and PLSTW347Af 5'- GTG TTT ACT GGG ACC ACA ACA GCG GCA GGT GGC G -3' (forward primer) and PLSTW347Ab 5'- C GCC ACC TGC CGC TGT TGT GGT CCC AGT AAA CAC -3' (backward primer) were designed for Trp347Gly and Trp347Ala mutants, respectively. The PCR used artificial 2,6SiaTple gene, which has been optimized for the expression in *E. coli*, in vector pET19b (pET19-PLST) as template.

PCR reaction:

| | |
|---|---|
| dsDNA template: | 2 $\mu$l |
| Primer 1: | 2 $\mu$l |
| Primer 2: | 2 $\mu$l |
| dNTP mix: | 1 $\mu$l |
| 10×Reaction Buffer: | 5 $\mu$l |
| QuikSolution reagent: | 1.5 $\mu$l |
| QuikChange Lightning Enzyme: | 1 $\mu$l |
| dd$H_2O$: | up to 50 $\mu$l |

| | | |
|---|---|---|
| 95°C | 2 min | |
| 95°C | 30 sec | |
| 72°C | 10 sec | |
| 68°C | 4 min | 30 cycles |
| 68°C | 10 min | |

The PCR product was digested by 2 $\mu$l *Dpn*I for 1 h at 37°C. Then 2 $\mu$l PCR products was transformed into Jm109 competent cells and cultured on ampicillin agar plates overnight at 37°C. The positive colonies were picked into 5 mL LB-ampicillin medium. After overnight culture at 37°C 250 rpm, the amplified plasmids pET19-PLSTW347G and pET19-PLSTW347A were extracted, purified by silica column and sent for sequencing.

The construction of Trp345 single-site mutagenesis library of 2,6SiaTple used QuikChange

Lighning Mutagenesis Kit. PCR primers PLSTTW345f 5'- TTT GTG TTT ACT GGG ACC NNS ACA GGG GCA GGT GGC GAG -3' (forward primer) and PLSTTW345b 5'- C GCC ACC TGC CCC SNN TGT GGT CCC AGT AAA CAC AAA ATT C -3' (backward primer); and PLSTAW345f 5'- TTT GTG TTT ACT GGG ACC NNS ACA GCG GCA GGT GGC GAG -3' (forward primer) and PLSTAW345b 5'- CTC GCC ACC TGC CGC TGT SNN GGT CCC AGT AAA CAC AAA -3' (backward primer) were designed for Trp345 site. The PCR used pET19-PLSTW347G and pET19-PLSTW347A plasmids as templates.

PCR reaction:

| | |
|---|---|
| dsDNA template: | 2 $\mu$l |
| Primer 1: | 2 $\mu$l |
| Primer 2: | 2 $\mu$l |
| dNTP mix: | 1 $\mu$l |
| 10×Reaction Buffer: | 5 $\mu$l |
| QuikSolution reagent: | 1.5 $\mu$l |
| QuikChange Lightning Enzyme: | 1 $\mu$l |
| $ddH_2O$: | up to 50 $\mu$l |

| | | |
|---|---|---|
| 95°C | 2 min | |
| 95°C | 30 sec | |
| 68°C | 4 min | 30 cycles |
| 68°C | 10 min | |

The PCR product was digested by 2 $\mu$l *Dpn*I for 1 h at 37°C. Then 2 $\mu$l PCR product was transformed into BL21(DE3) competent cells and cultured on ampicillin agar plates overnight at 37°C. 96 colonies were picked into 96-well plates containing 150 $\mu$l/well LB-ampicillin medium. After overnight culture at 37°C 900 rpm, 30 $\mu$l glycerol was added into each well. The whole plates were sealed with plastic lids and stored at -80°C.

### 13.7.3. Screening of mutagenesis libraries of SiaT

2 $\mu$l of each clone in libraries was transferred to individual wells in 96-well round bottom plates containing 150 $\mu$l/well LB growth media comprised 30 $\mu$g/mL Amp. The plates were sealed with plastic lids to avoid evaporation, and then incubated at 30°C, 1,000 rpm overnight. After that, 5 $\mu$l was transferred from each well to 96-deepwell plates with plastic lids

containing 400 $\mu$l/well of LB-amp growth medium containing 0.1 mM IPTG. The plates were then incubated at 30°C, 1,000 rpm overnight, until the amount of cells in each well reached the maximum. Cells were harvested by centrifugation at 4,000 rpm for 20 min. The culture medium was removed and the cell pellets could be stored at -80°C.

The cell pellets were resuspended in 150 $\mu$l resuspension buffer (1/10 BugBuster solution (Novagen, Germany), 0.5 mg/mL lysozyme (Roth, Germany) and 4U/mL Benzonase endonuclease (Novagen, Germany)) and incubated in a shaker for 30 min at room temperature. Then, the cell lysate was centrifugated at 4,000 rpm for 30 min. 40 $\mu$l of the supernatant was transferred to a new 96-well flat bottom plate. 140 $\mu$l of substrate solution (2 mM Tris-HCl (pH 8.0), 0.28 mM phenol red and 50 mM Gal-α-O-$CH_3$ was added to each well. The reaction was initiated by the addition of 1 mM CMP-Neu5Ac. The total volume in each well was 200 $\mu$l. The OD increase was read by the plate-reader at 560 nm.

20 positive colonies with the highest activity were picked from each plate into a new plate for the second round screening and so on.

## 13.8. Chemoenzymatic Synthesis of neo-2,3- and 2,6-Sialoconjugates

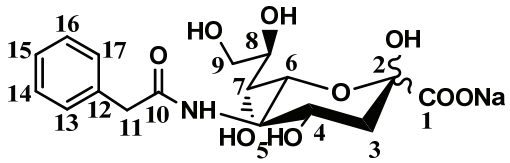
### 13.8.1. Enzymatic synthesis of sialic acids

#### 13.8.1.1. *N*-phenyacetyl-D-neuraminic acid (Neu5PhAc)

50 mg ManNPhAc (1 eq.) and 58 mg pyruvate (3 eq.) were added into a 50 ml flask with 50 mM PBS (pH 7.5). 5 mg aldolase was added to start the reaction for 12 h. When reaction was completed, 300 ml methanol was added. The flask was kept at -20°C for 30 min. Then the reaction mixture was filted to remove the precipitate. The supernatant was concentrated by rotary evaporator. The product was purified by Biogel P-2 with water as mobile phase. 40.9 mg ManNPhAc (white powder) was harvested with a yield of 59%.

***N*-phenylacetyl-D-neuraminic acid**

**(Neu5PhAc)**




Chemical Formula: $C_{17}H_{22}NNaO_9$
Molecular Weight: 407.35

$^{1}$H NMR (500 MHz, Deuterium Oxide) δ=7.43 – 7.38 (m, 2H, 13-, 17H), 7.37 – 7.32 (m, 3H, 14-, 15-, 16-H), 4.08 – 3.95 (m, 2H, 4-, 6-H), 3.89 (t, *J* = 10.2 Hz, 1H, 5-H), 3.74 (dd, *J* = 11.8, 2.8 Hz, 1H, 9a-H), 3.70 – 3.66 (m, 1H, 8-H), 3.64 (d, *J* = 6.9 Hz, 2H, 11-H), 3.48 (dd, *J* = 11.8, 6.5 Hz, 1H, 9b-H), 3.31 (dd, *J* = 9.0, 1.2 Hz, 1H, 7-H), 2.20 (dd, *J* = 13.0, 4.9 Hz, 1H, 3a-H), 1.81 (dd, *J* = 13.0, 11.6 Hz, 1H, 3b-H) ppm.

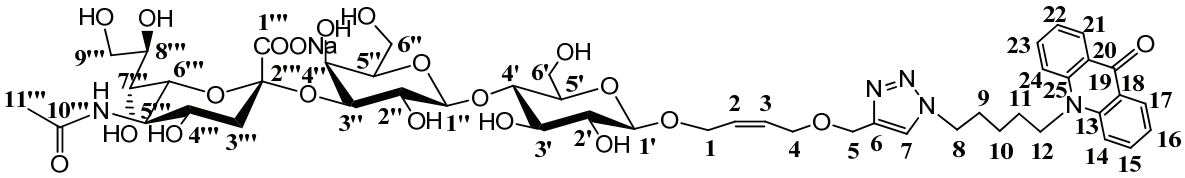
$^{13}$C NMR (126 MHz, Deuterium Oxide) δ=175.35 (C-1), 170.30 (C-10), 135.19 (C-12), 129.15 (C-15), 129.03 (C-14, -16), 127.39 (C-13, -17), 96.49 (C-2), 70.46 (C-6), 70.29 (C-8), 68.63 (C-7), 67.12 (C-4), 63.40 (C-9), 52.43 (C-5), 42.75 (C-11), 39.58 (C-3) ppm.

## 13.8.2. Enzymatic synthesis of neo-2,3-sialoconjugates with various donors

**10-[5-(4-(((Z)-4-*O*-((*N*-acetyl-α-D-neuraminic acid)-(2→3)-β-D-lactosyloxy)-but-2-enyloxy)methyl)-1*H*-1,2,3-triazol-1-yl)pentyl]-10*H*-acridin-9-on**

**(Neu5Ac-2,3-Lac-D-T-P-Acr)**

For the structure identification of the product catalyzed by recombinant 2,3SiaT$_{pph}$, 25.2 mg CTP (2 eq.) and 13.1 mg Neu5Ac (1.5 eq.) were added into 50 mL Tris-HCl (pH8.6) buffer with 0.5 M NaCl and 0.1% Triton ×100. 5 mg CSS and 20 μl PPase were added to start the reaction. When the CMP-activation was completed, 20 mg Lac-D-T-P-Acr (1 eq.) and 5 mg recombinant 2,3SiaT$_{pph}$ were added to start the reaction. The pH was being adjusted for keeping at 8.6 by pH-stat during the reaction for 2-3 h. When the reaction finished, 300 mL -20°C methanol was added to stop the reaction and precipitate the most of proteins in the reaction system. The precipitate was removed by filtration. The filtrate was concentrated to 10 mL and loaded on Biogel P-2 (Bio-Rad, Germany) column (3 cm × 100 cm). The products were further washed out by dd$H_2O$ and collected and lyophilized for NMR determination. Product: 23.7 mg yellow powder. Yield: 84%.



Chemical Formula: $C_{48}H_{64}N_5NaO_{21}$
Molecular Weight: 1070.03

$^1$H NMR (500 MHz, Deuterium Oxide) δ=7.78 (dd, *J* = 8.0, 1.6 Hz, 2H, 17-, 21-H), 7.73 (s, 1H, 7-H), 7.30 – 7.26 (m, 2H, 15-, 23-H), 6.93 – 6.77 (m, 4H, 24-, 14-, 16-, 22-H), 5.70 – 5.48 (m, 2H, 2-, 3-H), 4.45 (s, 2H, 5-H), 4.26 (d, *J* = 8.0 Hz, 1H, 1”-H), 4.23 – 4.06 (m, 4H, 1’-, 1a-, 8-H), 4.05 (dd, *J* = 9.9, 3.1 Hz, 1H, 1b-H), 3.95 – 3.90 (m, 3H, 3”-, 2-, 4-H), 3.87 – 3.75 (m, 5H, 4”-, 6”’-, 9”’a-, 6’-, 5”’-H), 3.73 – 3.46 (m, 13H, 6’b-, 6”-, 8”’-, 7”’-, 9”’b-, 3’-, 4”’-, 4’-, 2”-, 8-, 5”-H), 3.37 – 3.28 (m, 1H, 5’-H), 3.20 (dd, *J* = 9.2, 8.0 Hz, 1H, 2’-H), 2.72 (dd, *J*

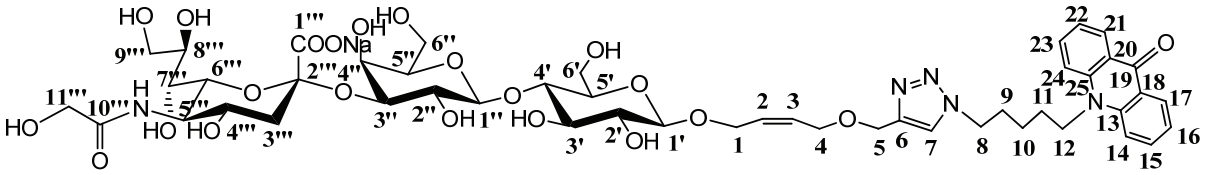
= 12.4, 4.7 Hz, 1H, 3'''a-H), 1.98 (s, 3H, 11'''-H), 1.75 (t, *J* = 12.1 Hz, 1H, 3'''b-H), 1.56 – 1.50 (m, 2H, 9-H), 1.15 – 0.87 (m, 4H, 11-, 10-H) ppm.

$^{13}$C NMR (126 MHz, Deuterium Oxide) δ=178.13 (C-19), 174.90 (C-1'''), 173.47 (C-10'''), 143.79 (C-6), 140.59 (C-13, -25), 134.57 (C-15, -23), 129.23 (C-2), 129.00 (C-3), 126.18 (C-17, -21), 124.65 (C-7), 121.66 (C-16, -22), 120.33 (C-18, -20), 115.02 (C-14, -24), 103.28 (C-1"), 100.92 (C-1'), 100.31 (C-2'''), 79.59 (C-4'), 74.63 (C-3"), 73.69 (C-8'''), 72.63 (C-5'), 72.55 (C-5"), 72.42 (C-3'), 71.80 (C-2'), 70.79 (C-6'''), 68.52 (C-2"), 68.40 (C-7'''), 68.36 (C-4'''), 65.05 (C-4), 64.63 (C-1), 63.50 (C-9'''), 62.67 (C-5), 62.27 (C-6"), 60.22 (C-6'), 51.83 (C-5'''), 50.05 (C-8), 45.19 (C-12), 40.14 (C-3'''), 28.77 (C-9), 25.73 (C-11), 22.60 (C-10), 22.09 (C-11''') ppm.

**10-[5-(4-(((Z)-4-*O*-((*N*-glycolyl-α-D-neuraminic acid)-(2→3)-β-D-lactosyloxy)-but-2-enyloxy)methyl)-1*H*-1,2,3-triazol-1-yl)pentyl]-10*H*-acridin-9-on**

**(Neu5Gc-2,3-Lac-D-T-P-Acr)**

For the structure identification of the product catalyzed by recombinant 2,3SiaT$_{pph}$, 25.2 mg CTP (2 eq.) and 13.7 mg Neu5Gc (1.5 eq.) were added into 50 mL Tris-HCl (pH8.6) buffer with 0.5 M NaCl and 0.1% Triton ×100. 5 mg CSS and 20 μl PPase were added to start the reaction. When the CMP-activation was completed, 20 mg Lac-D-T-P-Acr (1 eq.) and 5 mg recombinant 2,3SiaT$_{pph}$ were added to start the reaction. The pH was being adjusted for keeping at 8.6 by pH-stat during the reaction for 2-3 h. When the reaction finished, 300 mL -20°C methanol was added to stop the reaction and precipitate the most of proteins in the reaction system. The precipitate was removed by filtration. The filtrate was concentrated to 10 mL and loaded on Biogel P-2 (Bio-Rad, Germany) column (3 cm × 100 cm). The products were further washed out by dd$H_2O$ and collected and lyophilized for NMR determination. Product: 22.1 mg yellow powder. Yield: 77%.



Chemical Formula: $C_{48}H_{64}N_5NaO_{22}$
Molecular Weight: 1086.03

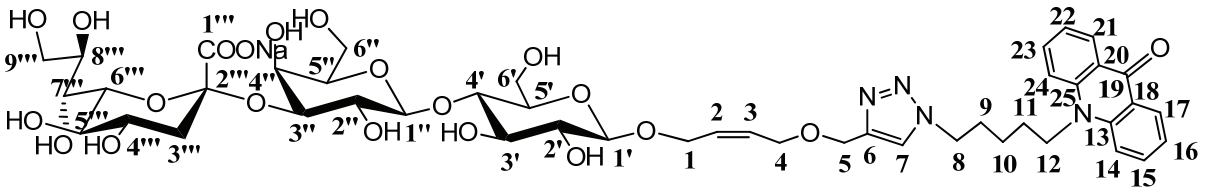
$^{1}$H NMR (500 MHz, Deuterium Oxide) δ=7.72 (dd, *J* = 8.1, 1.3 Hz, 2H, 17-, 21-H), 7.64 (s, 1H, 7-H), 7.30 – 7.18 (m, 2H, 15-, 23-H), 6.89 – 6.76 (m, 4H, 24-, 14-, 16-, 22-H), 5.62 – 5.45 (m, 2H, 2-, 3-H), 4.36 (s, 2H, 5-H), 4.32 (d, *J* = 7.9 Hz, 1H, 1”-H), 4.15 (d, *J* = 8.0 Hz, 1H, 1’-H), 4.13-4.04 (m, 3H, 1a-, 8-H), 4.02 – 3.93 (m, 4H, 11’”-, 1b-, 3”-H), 3.87 – 3.79 (m, 4H, 4-, 4”-, 6’”-, 9’”a-H ), 3.79 – 3.57 (m, 12H, 6’a-, 5’”-, 6’b-, 6”-, 8’”-, 7’”-, 9’”b-, 3’-, 4’”-, 4’,-2”-H), 3.57 – 3.46 (m, 3H, 8-, 5”-H), 3.28 – 3.20 (m, 1H, 5’-H), 3.09 (dd, *J* = 9.2, 8.1 Hz, 1H, 2’-H), 2.69 – 2.59 (m, 2H, 3’”a-H), 1.66 (t, *J* = 12.1 Hz, 1H, 3’”b-H), 1.55-1.35 (m, 2H, 9-H), 1.09 – 0.80 (m, 4H, 11-, 10-H) ppm.

$^{13}$C NMR (126 MHz, Deuterium Oxide) δ=178.17 (C-19), 175.82 (C-1’”), 173.95 (C-10’”), 143.87 (C-6), 140.65 (C-13, -25), 134.60 (C-15, -23), 129.28 (C-2), 129.05 (C-3), 126.25 (C-17, -21), 124.66 (C-7), 121.68 (C-16, -22), 120.41 (C-18, -20), 115.05 (C-14, -24), 102.80 (C-1”), 101.13 (C-1’), 99.89 (C-2’”), 78.42 (C-4’), 75.60 (C-3”), 75.19 (C-8’”), 74.80 (C-5’), 74.45 (C-5”), 72.74 (C-3’), 72.70 (C-2’), 71.86 (C-6’”), 69.41 (C-2”), 68.13 (C-7’”), 67.53 (C-4’”), 65.10 (C-4), 64.73 (C-1), 62.65 (C-9’”), 62.35 (C-5), 61.08 (C-6”, -11’”), 60.10 (C-6’), 51.51 (C-5’”), 50.09 (C-8), 45.24 (C-12), 39.81 (C-3’”), 28.81 (C-9), 25.79 (C-11), 22.66 (C-10) ppm.

**10-[5-(4-(((Z)-4-*O*-((3-deoxy-β-D-*glycero*-D-*galacto*-2-nonulopyronosonic acid)-(2→3)-β-D-lactosyloxy)-but-2-enyloxy)methyl)-1*H*-1,2,3-triazol-1-yl)pentyl]-10*H*-acridin-9-on**

**(Kdn-2,3-Lac-D-T-P-Acr)**

For the structure identification of the product catalyzed by recombinant 2,3SiaT$_{pph}$, 25.2 mg CTP (2 eq.) and 11.5 mg Kdn (1.5 eq.) were added into 50 mL Tris-HCl (pH8.6) buffer with 0.5 M NaCl and 0.1% Triton ×100. 5 mg CSS and 20 μl PPase were added to start the reaction. When the CMP-activation was completed, 20 mg Lac-D-T-P-Acr (1 eq.) and 5 mg recombinant 2,3SiaT$_{pph}$ were added to start the reaction. The pH was being adjusted for keeping at 8.6 by pH-stat during the reaction for 2-3 h. When the reaction finished, 300 mL -20°C methanol was added to stop the reaction and precipitate the most of proteins in the reaction system. The precipitate was removed by filtration. The filtrate was concentrated to 10 mL and loaded on Biogel P-2 (Bio-Rad, Germany) column (3 cm × 100 cm). The products were further washed out by dd$H_2O$ and collected and lyophilized for NMR determination. Product: 23.4 mg yellow powder. Yield: 86%.



Chemical Formula: $C_{46}H_{61}N_4NaO_{21}$
Molecular Weight: 1028.98

$^1$H NMR (500 MHz, Deuterium Oxide) δ=7.83 (dd, *J* = 8.0, 1.6 Hz, 2H, 17-, 21-H), 7.76 (s, 1H, 7-H), 7.38 – 7.26 (m, 2H, 15-, 23-H), 6.97 – 6.84 (m, 4H, 24-, 14-, 16-, 22-H), 5.75 – 5.54 (m, 2H, 2-, 3-H), 4.48 (s, 2H, 5-H), 4.29 (d, *J* = 8.0 Hz, 1H, 1”-H), 4.26 – 4.10 (m, 4H, 1’-, 1a-, 8-H), 4.07 (dd, *J* = 9.8, 2.7 Hz, 1H, 1b-H), 4.00-3.92 (m, 3H, 3”-, 2-, 4-H), 3.92 – 3.82 (m, 4H, 4”-, 6”’-, 9”’a-, 6’-H), 3.77 – 3.45 (m, 14H, 5”’-, 6’b-, 6”-, 8”’-, 7”’-, 9”’b-, 3’-, 4”’-, 4’-, 2”-, 8-, 5”-H), 3.39-3.34 (m, 1H, 5’-H), 3.24 (dd, *J* = 9.2, 8.0 Hz, 1H, 2’-H), 2.70 (dd, *J* = 12.5, 4.6 Hz, 1H, 3”’a-H), 1.74 (t, *J* = 12.0 Hz, 1H, 3”’b-H), 1.79 – 1.69 (m, 2H, 9-H), 1.18 – 0.92 (m, 4H, 11-, 10-H) ppm.

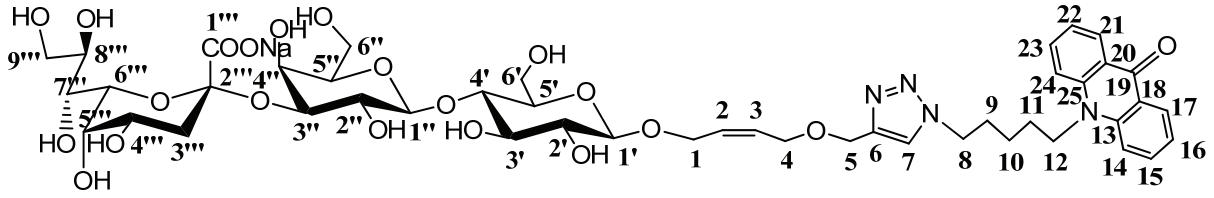
$^{13}$C NMR (126 MHz, Deuterium Oxide) δ=178.05 (C-19), 174.11 (C-1”’), 143.88 (C-6), 140.61 (C-13, -25), 134.53 (C.15, -23), 129.29 (C-2), 129.04 (C-3), 126.25 (C-17, -21), 124.62 (C-7), 121.61 (C-16, -22), 120.43 (C-18, -20), 114.97 (C-14, -24), 102.82 (C-1”), 101.17 (C-1’), 99.87 (C-2”’), 78.43 (C-4’), 75.56 (C-3”), 75.20 (C-8”’), 74.79 (C-5’), 74.46 (C-5”), 73.99 (C-3’), 72.75 (C-2’), 72.14 (C-6”’), 70.34 (C-2”), 69.85 (C-7”’), 69.40 (C-4”’), 67.83 (C-5”’), 65.17 (C-4), 64.76 (C-1), 62.76 (C-9”’), 62.37 (C-5), 61.06 (C-6”), 60.11 (C-6’), 50.05 (C-8), 45.24 (C-12), 39.39 (C-3”’), 28.83 (C-9), 25.78 (C-11), 22.66 (C-10) ppm.

**10-[5-(4-(((Z)-4-*O*-((3-deoxy-α-D-*glycero*-D-*gulo*-2-nonulopyranosonic acid)-(2→3)-β-D-lactosyloxy)-but-2-enyloxy)methyl)-1*H*-1,2,3-triazol-1-yl)pentyl]-10*H*-acridin-9-on**

**(*epi*-Kdn-2,3-Lac-D-T-P-Acr)**

For the structure identification of the product catalyzed by recombinant 2,3SiaT$_{pph}$, 25.2 mg CTP (2 eq.) and 11.5 mg *epi*-Kdn (1.5 eq.) were added into 50 mL Tris-HCl (pH8.6) buffer with 0.5 M NaCl and 0.1% Triton ×100. 5 mg CSS and 20 μl PPase were added to start the reaction. When the CMP-activation was completed, 20 mg Lac-D-T-P-Acr (1 eq.) and 5 mg

recombinant 2,3SiaT$_{pph}$ were added to start the reaction. The pH was being adjusted for keeping at 8.6 by pH-stat during the reaction for 2-3 h. When the reaction finished, 300 mL -20°C methanol was added to stop the reaction and precipitate the most of proteins in the reaction system. The precipitate was removed by filtration. The filtrate was concentrated to 10 mL and loaded on Biogel P-2 (Bio-Rad, Germany) column (3 cm × 100 cm). The products were further washed out by dd$H_2O$ and collected and lyophilized for NMR determination. Product: 15.8 mg yellow powder. Yield: 58%.



Chemical Formula: $C_{46}H_{61}N_4NaO_{21}$
Molecular Weight: 1028.98

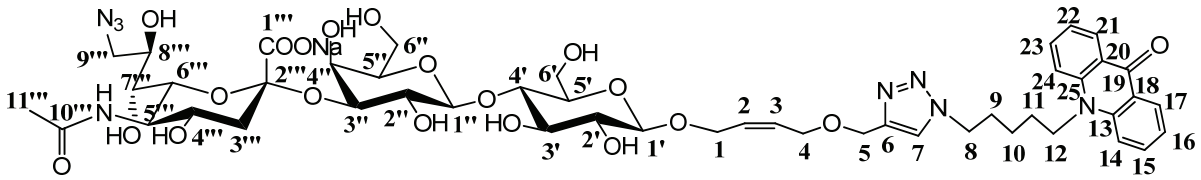
$^1$H NMR (500 MHz, Deuterium Oxide) δ=7.75 – 7.67 (m, 2H, 17-, 21-H), 7.62 (s, 1H, 7-H), 7.29 – 7.14 (m, 2H, 15-, 23-H), 6.89 – 6.68 (m, 4H, 24-, 14-, 16-, 22-H), 5.62 – 5.39 (m, 2H, 2-, 3-H), 4.35 (s, 2H, 5-H), 4.15 (d, *J* = 8.0 Hz, 1H, 1”-H), 4.12 – 3.96 (m, 5H, 1’-, 1a-, 8-, 1b-H), 3.86 – 3.67 (m, 3H, 3”-, 2-, 4-, 4”-, 6”’-, 9”’a-, 6’-, 5”’-H), 3.64 – 3.33 (m, 13H, 6’b-, 6”-, 8”’-, 7”’-, 9”’b-, 3’-, 4”’-, 4’-, 2”-, 8-, 5”-H), 3.26 – 3.20 (m, 1H, 5’-H), 3.10 (dd, *J* = 9.3, 8.0 Hz, 1H, 2’-H), 2.34 (dd, *J* = 12.2, 4.6 Hz, 1H, 3”’a-H), 1.82 (t, *J* = 12.4 Hz, 1H, 3”’b-H), 1.50 – 1.38 (m, 2H, 9-H), 1.03 – 0.83 (m, 4H, 11-, 10-H) ppm.

$^{13}$C NMR (126 MHz, Deuterium Oxide) δ=178.12 (C-19), 174.22 (C-1”’), 143.88 (C-6), 140.64 (C-13, -25), 134.56 (C.15, -23), 129.27 (C-2), 129.03 (C-3), 126.26 (C-17, -21), 124.63 (C-7), 121.65 (C-16, -22), 120.42 (C-18, -20), 115.02 (C-14, -24), 102.71 (C-1”), 101.16 (C-1’), 100.13 (C-2”’), 78.33 (C-4’), 75.54 (C-3”), 75.22 (C-8”’), 74.80 (C-5’), 74.44 (C-5”), 73.02 (C-3’), 72.75 (C-2’), 71.88 (C-6”’), 71.68 (C-2”), 69.55 (C-7”’), 69.14 (C-4”’), 67.47 (C-5”’), 65.13 (C-4), 64.75 (C-1), 62.35 (C-9”’), 62.18 (C-5), 61.08 (C-6”), 60.12 (C-6’), 50.07 (C-8), 45.23 (C-12), 34.09 (C-3”’), 28.82 (C-9), 25.79 (C-11), 22.66 (C-10) ppm.

**10-[5-(4-(((Z)-4-*O*-((9-aido-*N*-acetyl-α-D-neuraminic acid)-(2→3)-β-D-lactosyloxy)-but-2-enyloxy)methyl)-1*H*-1,2,3-triazol-1-yl)pentyl]-10*H*-acridin-9-on**

**(Neu5Ac9$N_3$-2,3-Lac-D-T-P-Acr)**

For the structure identification of the product catalyzed by recombinant 2,3SiaT$_{pph}$, 25.2 mg CTP (2 eq.) and 14.1 mg Neu5Ac9$N_3$ (1.5 eq.) were added into 50 mL Tris-HCl (pH8.6) buffer with 0.5 M NaCl and 0.1% Triton ×100. 5 mg CSS and 20 μl PPase were added to start the reaction. When the CMP-activation was completed, 20 mg Lac-D-T-P-Acr (1 eq.) and 5 mg recombinant 2,3SiaT$_{pph}$ were added to start the reaction. The pH was being adjusted for keeping at 8.6 by pH-stat during the reaction for 2-3 h. When the reaction finished, 300 mL -20°C methanol was added to stop the reaction and precipitate the most of proteins in the reaction system. The precipitate was removed by filtration. The filtrate was concentrated to 10 mL and loaded on Biogel P-2 (Bio-Rad, Germany) column (3 cm × 100 cm). The products were further washed out by dd$H_2O$ and collected and lyophilized for NMR determination. Product: 23.7 mg yellow powder. Yield: 82%.



Chemical Formula: $C_{48}H_{63}N_8NaO_{20}$
Molecular Weight: 1095.05

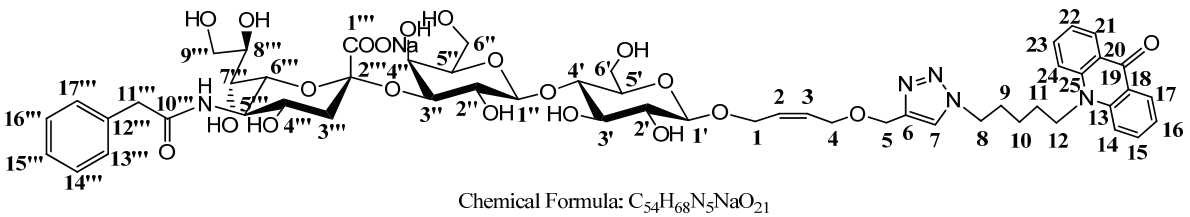
$^1$H NMR (500 MHz, Deuterium Oxide) δ=7.80 – 7.69 (m, 2H, 17-, 21-H), 7.65 (s, 1H, 7-H), 7.31 – 7.08 (m, 2H, 15-, 23-H), 6.91 – 6.60 (m, 4H, 24-, 14-, 16-, 22-H), 5.65 – 5.41 (m, 2H, 2-, 3-H), 4.38 (s, 2H, 5-H), 4.20 (d, *J* = 7.9 Hz, 1H, 1”-H), 4.17 – 4.00 (m, 4H, 1’-, 1a-, 8-H), 3.95 (dd, *J* = 9.8, 3.2 Hz, 1H, 1b-H), 3.92 – 3.84 (m, 3H, 3”-, 2-, 4-H), 3.84 – 3.68 (m, 4H, 4”-, 6”’-, 9”’a-, 6’-H), 3.68 – 3.35 (m, 14H, 5”’-, 6’b-, 6”-, 8”’-, 7”’-, 9”’b-, 3’-, 4”’-, 4’-, 2”-, 8-, 5”-H), 3.28 – 3.23 (m, 1H, 5’-H), 3.13 (dd, *J* = 9.1, 8.0 Hz, 1H, 2’-H), 2.65 (dd, *J* = 12.4, 4.6 Hz, 1H, 3”’a-H), 1.92 (s, 3H, 11”’-H), 1.67 (t, J = 12.1 Hz, 1H, 3”’b-H), 1.52 – 1.35 (m, 2H, 9-H), 1.08 – 0.81 (m, 4H, 11-, 10-H) ppm.

$^{13}$C NMR (126 MHz, Deuterium Oxide) δ=178.03 (C-19), 175.05 (C-1'''), 173.80 (C-10'''), 143.88 (C-6), 140.62 (C-13, -25), 134.51 (C-15, -23), 129.30 (C-2), 129.04 (C-3), 126.29 (C-17, -21), 124.59 (C-7), 121.59 (C-16, -22), 120.47 (C-18, -20), 114.95 (C-14, -24), 102.80 (C-1"), 101.12 (C-1'), 99.83 (C-2'''), 78.36 (C-4'), 75.75 (C-3"), 75.21 (C-8'''), 74.82 (C-5'), 74.45 (C-5"), 72.78 (C-3'), 70.44 (C-2'), 69.36 (C-6'''), 68.94 (C-2"), 68.30 (C-7'''), 67.46 (C-4'''), 65.17 (C-4), 64.69 (C-1), 62.38 (C-5), 61.04 (C-6'''), 60.16 (C-6'), 53.24 (C-9'''), 51.82 (C-5'''), 50.03 (C-8), 45.25 (C-12), 39.92 (C-3'''), 28.86 (C-9), 25.79 (C-11), 22.68 (C-10), 22.14 (C-11''') ppm.

**10-[5-(4-(((Z)-4-*O*-((*N*- phenylacetyl -α-D-neuraminic acid)-(2→3)-β-D-lactosyloxy)-but-2-enyloxy)methyl)-1*H*-1,2,3-triazol-1-yl)pentyl]-10*H*-acridin-9-on**

**(Neu5NPhAc-2,3-Lac-D-T-P-Acr)**

For the structure identification of the product catalyzed by recombinant 2,3SiaT$_{pph}$, 25.2 mg CTP (2 eq.) and 16.1 mg Neu5NPhAc (1.5 eq.) were added into 50 mL Tris-HCl (pH8.6) buffer with 0.5 M NaCl and 0.1% Triton ×100. 5 mg CSS-F192A and 20 μl PPase were added to start the reaction. When the CMP-activation was completed, 20 mg Lac-D-T-P-Acr (1 eq.) and 5 mg recombinant 2,3SiaT$_{pph}$ were added to start the reaction. The pH was being adjusted for keeping at 8.6 by pH-stat during the reaction for 2-3 h. When the reaction finished, 300 mL -20°C methanol was added to stop the reaction and precipitate the most of proteins in the reaction system. The precipitate was removed by filtration. The filtrate was concentrated to 10 mL and loaded on Biogel P-2 (Bio-Rad, Germany) column (3 cm × 100 cm). The products were further washed out by dd$H_2O$ and collected and lyophilized for NMR determination. Product: 14.5 mg yellow powder. Yield: 48%.



Chemical Formula: $C_{54}H_{68}N_5NaO_{21}$
Molecular Weight: 1146.13

$^{1}$H NMR (500 MHz, Deuterium Oxide) δ=7.75 (dd, *J* = 8.1, 1.7 Hz, 2H, 17-, 21-H), 7.61 (s, 1H, 7-H), 7.31 – 7.20 (m, 2H, 13'''-, 17'''-H), 7.19 – 7.05 (m, 5H, 15'''-, 14'''-, 16'''-, 15-, 23-H), 6.92 – 6.69 (m, 4H, 24-, 14-, 16-, 22-H), 5.63 – 5.37 (m, 2H, 2-, 3-H), 4.34 (s, 2H, 5-H), 4.14 (d, *J* = 7.9 Hz, 1H, 1"-H), 4.10 – 3.88 (m, 5H, 1'-, 1a-, 8-, 1b-H), 3.84 – 3.74 (m, 3H, 3"-, 2-, 4-H), 3.72 – 3.55 (m, 9H, 4"-, 6'''-, 9'''a-, 6'-, 5'''-, 11'''-, 6"-H), 3.54 – 3.30 (m, 11H, 6'b-, 8'''-, 7'''-, 9'''b-, 3'-, 4'''-, 4'-, 2"-, 8-, 5"-H), 3.24 – 3.19 (m, 1H, 5'-H), 3.09 (dd, *J* = 9.2, 8.0 Hz, 1H, 2'-H), 2.68 – 2.55 (m, 1H, 3'''a-H), 1.63 (t, *J* = 12.0 Hz, 1H, 3'''b-H), 1.52 – 1.35 (m, 2H, 9-H), 1.11 – 0.77 (m, 4H, 11-, 10-H) ppm.

$^{13}$C NMR (126 MHz, Deuterium Oxide) δ 178.21 (C-19), 175.60 (C-1'''), 173.88 (C-10'''), 143.85 (C-6), 140.71 (C-13, -25), 135.04 (C-15, -23), 134.61 (C-12'''), 129.34 (C-14'''), 129.09 (C-2), 129.01 (C-3), 128.91 (C-13''', -15'''), 127.30 (C-12''', -16'''), 126.29 (C-17, -21), 124.64 (C-7), 121.69 (C-16, -22), 120.48 (C-18, -20), 115.08 (C-14, -24), 102.82 (C-1"), 100.99 (C-1'), 99.81 (C-2'''), 78.35 (C-4'), 75.62 (C-3"), 75.21 (C-8'''), 74.79 (C-5'), 74.44 (C-5"), 72.99 (C-3'), 72.74 (C-2'), 71.85 (C-6'''), 69.37 (C-2"), 68.32 (C-7'''), 67.53 (C-4'''), 65.08 (C-4), 64.60 (C-1), 62.80 (C-9'''), 62.35 (C-5), 61.05 (C-6"), 60.03 (C-6'), 59.84 (C-12), 51.92 (C-5'''), 50.08 (C-8), 42.60 (C-11'''), 39.95 (C-3'''),28.84 (C-9), 25.82 (C-11), 22.69 (C-10) ppm.

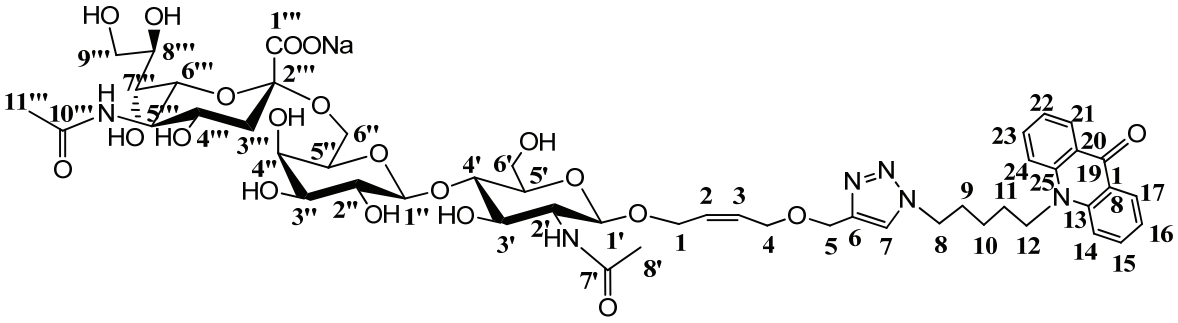
### 13.8.3. Enzymic synthesis of neo-2,6-sialoconjugates with various donors

**10-[5-(4-(((Z)-4-*O*-((*N*-acetyl-α-D-neuraminic acid)-(2→6)-β-D-galactopyranosyl-(1,4)-2-acetamido-2-deoxy-β-D-glucopyranosyloxy)-but-2-enyloxy)methyl)-1*H*-1,2,3-triazol-1-yl)pentyl]-10*H*-acridin-9-on**

**(Neu5Ac-2,6-LacNAc-D-T-P-Acr)**

For the structure identification of the product catalyzed by recombinant 2,6SiaT$_{ple}$, 25.2 mg CTP (2 eq.) and 12.5 mg Neu5Ac (1.5 eq.) were added into 50 mL Tris-HCl (pH8.6) buffer with 0.5 M NaCl and 0.1% Triton ×100. 5 mg CSS and 20 μl PPase were added to start the reaction. When the CMP-activation was completed, 20 mg LacNAc-D-T-P-Acr (1 eq.) and 5 mg recombinant 2,6SiaT$_{ple}$ were added to start the reaction. The pH was being adjusted for keeping at 8.6 by pH-stat during the reaction for 2-3 h. When the reaction finished, 300 mL -

20°C methanol was added to stop the reaction and precipitate the most of proteins in the reaction system. The precipitate was removed by filtration. The filtrate was concentrated to 10 mL and loaded on Biogel P-2 (Bio-Rad, Germany) column (3 cm × 100 cm). The products were further washed out by dd$H_2O$ and collected and lyophilized for NMR determination. Product: 22.8 mg yellow powder. Yield: 82%.



Chemical Formula: $C_{50}H_{67}N_6NaO_{21}$
Molecular Weight: 1111.08

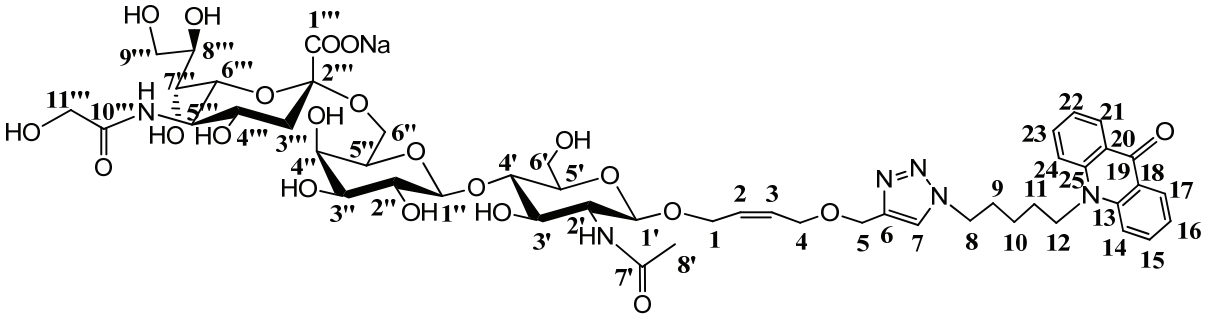
$^1$H NMR (500 MHz, Deuterium Oxide) δ=7.80 – 7.68 (m, 2H, 17-, 21-H), 7.63 (s, 1H, 7-H), 7.23 (ddd, *J* = 8.6, 6.7, 1.6 Hz, 2H, 15-, 23-H), 6.89 – 6.74 (m, 4H, 14-, 24-, 16-, 22-H), 5.56 – 5.42 (m, 2H, 2-, 3-H), 4.35 (s, 2H, 5-H), 4.24 (d, *J* = 7.9 Hz, 1H, 1”-H), 4.14 – 3.92 (m, 4H, 1’, 8-, 1a-H), 3.88 – 3.57 (m, 10H, 1b-, 4-, 9’”a-, 5’”-, 8’”-, 6’”-, 6’a-, 6”a-, 4”-H), 3.57 – 3.52 (m, 3H, 4’”-, 5”-, 6’b-H), 3.51 – 3.34 (m, 9H, 6”b-, 3”-, 9’”-, 12-, 7’”-, 3’-, 4’-, 2”-H), 3.34 – 0.28 (m, 1H, 5’-H), 2.52 (dd, *J* = 12.4, 4.6 Hz, 1H, 3’”a-H), 1.87 (s, 3H, 8’-H), 1.85 (s, 2H, 11’”-H), 1.55 (t, *J* = 12.1 Hz, 1H, 3’”-H), 1.51 - 1.39 (m, 2H, 9-H), 1.08 – 0.80 (m, 4H, 11-, 10-H) ppm.

$^{13}$C NMR (126 MHz, Deuterium Oxide) δ=178.16 (C-19), 174.95 (C-1’”), 174.34 (C-7’”), 173.53 (C-10’”), 143.85 (C-6), 140.65 (C-13, -25), 134.59 (C-15, -23), 129.51 (C-2), 128.88 (C-3), 126.22 (C-17, -21), 124.68 (C-7), 121.68 (C-16, -21), 120.40 (C-18, -20), 115.06 (C-14, -24), 103.56 (C-1”), 100.24 (C-1’), 99.71 (C-2’”), 80.72 (C-4’), 74.58 (C-3”), 73.73 (C-8’”), 72.63 (C-5’), 72.58 (C-5”), 72.54 (C-3’), 71.77 (C-2’), 70.79 (C-6’”), 68.50 (C-7’”), 68.48 (C-4’”), 68.27 (C-4”), 65.13 (C-4), 64.52 (C-1), 63.37 (C-6”), 62.75 (C-9’”), 62.30 (C-5), 60.40 (C-6’), 54.84 (C-2”), 52.00 (C-5’”), 50.08 (C-8), 45.22 (C-12), 40.18 (C-3’”), 28.82 (C-9), 25.78 (C-11), 22.65 (C-10), 22.37 (C-8’), 22.12 (C-11’”) ppm.

**10-[5-(4-(((Z)-4-*O*-((*N*-glycolyl-α-D-neuraminic acid)-(2→6)-β-D-galactopyranosyl-(1,4)-2-acetamido-2-deoxy-β-D-glucopyranosyloxy)-but-2-enyloxy)methyl)-1*H*-1,2,3-triazol-1-yl)pentyl]-10*H*-acridin-9-on**

**(Neu5Gc- 2,6-LacNAc-D-T-P-Acr)**

For the structure identification of the product catalyzed by recombinant 2,6SiaT$_{ple}$, 25.2 mg CTP (2 eq.) and 13 mg Neu5Gc (1.5 eq.) were added into 50 mL Tris-HCl (pH8.6) buffer with 0.5 M NaCl and 0.1% Triton×100. 5 mg CSS and 20 $\mu$l PPase were added to start the reaction. When the CMP-activation was completed, 20 mg LacNAc-D-T-P-Acr (1 eq.) and 5 mg recombinant 2,6SiaT$_{ple}$ were added to start the reaction. The pH was being adjusted for keeping at 8.6 by pH-stat during the reaction for 2-3 h. When the reaction finished, 300 mL -20°C methanol was added to stop the reaction and precipitate the most of proteins in the reaction system. The precipitate was removed by filtration. The filtrate was concentrated to 10 mL and loaded on Biogel P-2 (Bio-Rad, Germany) column (3 cm × 100 cm). The products were further washed out by dd$H_2O$ and collected and lyophilized for NMR determination. Product: 21.1 mg yellow powder. Yield: 75%.



Chemical Formula: $C_{50}H_{67}N_6NaO_{22}$
Molecular Weight: 1127.08

$^1$H NMR (500 MHz, Deuterium Oxide) δ=7.76 – 7.68 (m, 2H, 17-, 21-H), 7.63 (s, 1H, 7-H), 7.24 (ddd, *J* = 8.6, 6.8, 1.7 Hz, 2H, 15-, 23-H), 6.88 – 6.76 (m, 4H, 14-, 24-, 16-, 22-H), 5.55 – 5.43 (m, 2H, 2-, 3-H), 4.35 (s, 2H, 5-H), 4.24 (d, *J* = 7.9 Hz, 1H, 1”-H), 4.13 – 3.97 (m, 4H, 1’, 8-, 1a-H), 3.96 (s, 2H, 11’”H), 3.88 – 3.66 (m, 10H, 1b-, 4-, 9’”a-, 5’”-, 8’”-, 6’”-, 6’a-, 6”a-, 4”-H), 3.67 – 3.56 (s, m, 3H, 4’”-, 5”-, 6’b-H), 3.57 – 3.35 (m, 9H, 6”b-, 3”-, 9’”-, 12-, 7’”-, 3’-, 4’-, 2”-H), 3.34 – 3.27 (m, 1H, 5’-H), 2.54 (dd, *J* = 12.4, 4.7 Hz, 1H, 3’”a-H), 1.86 (s, 3H, 8’-

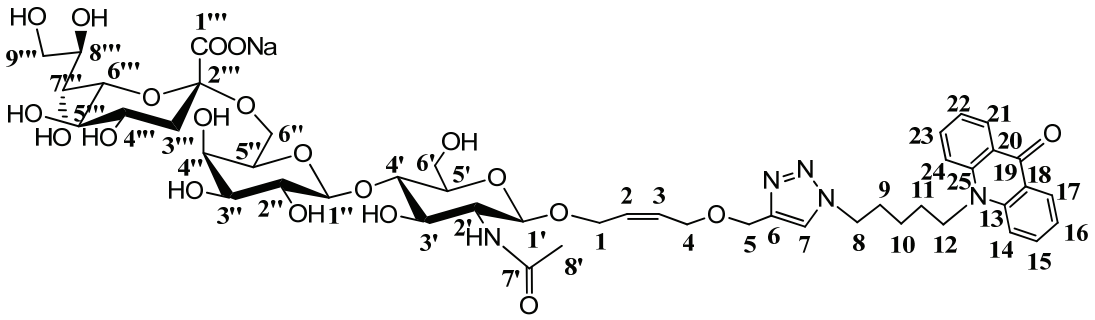
H), 1.56 (t, *J* = 12.1 Hz, 4, 4.6 Hz, 1H, 3'''a-H), 1.87 (s, 3H, 8'-H), 1.85 (s, 2H, 11'''-H), 1.55 (t, J = 12.1 Hz, 1H), 1.51 – 1.39 (m, 2H, 9-H), 1.05 – 0.82 (m, 4H, 11-, 10-H) ppm.

$^{13}$C NMR (126 MHz, Deuterium Oxide) δ=178.18 (C-19), 175.70 (C-1'''), 174.37 (C-7'''), 173.55 (C-10'''), 143.86 (C-6), 140.66 (C-13, -25), 134.60 (C-15, -23), 129.52 (C-2), 128.89 (C-3), 126.23 (C-17, -21), 124.69 (C-7), 121.69 (C-16, -21), 120.40 (C-18, -20), 115.08 (C-14, -24), 103.56 (C-1''), 100.28 (C-1'), 99.72 (C-2'''), 80.68 (C-4'), 74.59 (C-3''), 73.74 (C-8'''), 72.57 (C-5'), 72.54 (C-5''), 72.36 (C-3'), 71.83 (C-2'), 70.80 (C-6'''), 68.47 (C-7'''), 68.44 (C-4'''), 68.03 (C-4''), 65.12 (C-4), 64.53 (C-1), 63.40 (C-6''), 62.73 (C-9'''), 62.30 (C-5), 61.10 (C-11'''), 60.40 (C-6'), 54.86 (C-2''), 51.69 (C-5'''), 50.09 (C-8), 45.23 (C-12), 40.23 (C-3'''), 28.82 (C-9), 25.79 (C-11), 22.65 (C-10), 22.39 (C-8') ppm.

**10-[5-(4-(((Z)-4-*O*-((3-deoxy-β-D-*glycero*-D-*galacto*-2-nonulopyronosonic acid)-(2→6)-β-D- galactopyranosyl-(1,4)-2-acetamido-2-deoxy-β-D-glucopyranosyloxy)-but-2-enyloxy)methyl)-1*H*-1,2,3-triazol-1-yl)pentyl]-10*H*-acridin-9-on**

**(Kdn-2,6-LacNAc-D-T-P-Acr)**

For the structure identification of the product catalyzed by recombinant 2,6SiaT$_{ple}$, 25.2 mg CTP (2 eq.) and 11 mg Kdn (1.5 eq.) were added into 50 mL Tris-HCl (pH8.6) buffer with 0.5 M NaCl and 0.1% Triton ×100. 5 mg CSS and 20 μl PPase were added to start the reaction. When the CMP-activation was completed, 20 mg LacNAc-D-T-P-Acr (1 eq.) and 5 mg recombinant 2,6SiaT$_{ple}$ were added to start the reaction. The pH was being adjusted for keeping at 8.6 by pH-stat during the reaction for 2-3 h. When the reaction finished, 300 mL -20°C methanol was added to stop the reaction and precipitate the most of proteins in the reaction system. The precipitate was removed by filtration. The filtrate was concentrated to 10 mL and loaded on Biogel P-2 (Bio-Rad, Germany) column (3 cm × 100 cm). The products were further washed out by dd$H_2O$ and collected and lyophilized for NMR determination. Product: 22.1 mg yellow powder. Yield: 82%.



Chemical Formula: $C_{48}H_{64}N_5NaO_{21}$
Molecular Weight: 1070.03

$^1$H NMR (500 MHz, Deuterium Oxide) δ=7.85 (dd, *J* = 8.1, 1.7 Hz, 2H, 17- 21H), 7.76 (s, 1H, 7-H), 7.35 (ddd, *J* = 8.6, 6.8, 1.6 Hz, 2H, 15-, 23-H), 7.00 – 6.87 (m, 4H, 14-, 24-, 16-, 22-H), 5.70 – 5.53 (m, 2H, 2-, 3-H), 4.47 (s, 2H, 5-H), 4.37 (d, *J* = 7.9 Hz, 1H, 1”-H), 4.27 – 4.04 (m, 4H, 1’, 8-, 1a-H), 4.03 – 3.72 (m, 10H, 1b-, 4-, 9’”a-, 5’”-, 8’”-, 6’”-, 6’a-, 6”a-, 4”-H), 3.72 – 3.48 (m, 12H, 4’”-, 5”-, 6’b-, 6”b-, 3”-, 9’”-, 12-, 7’”-, 3’-, 4’-, 2”-H), 3.47 – 3.42 (m, 1H, 5’-H), 2.61 (dd, *J* = 12.4, 4.7 Hz, 1H, 3’”a-H), 2.00 (s, 3H, 8’-H), 1.64 (t, *J* = 12.2 Hz, 1H, 3’”-H), 1.62 – 1.53 (m, 2H, 9-H), 1.20 – 0.94 (m, 4H, 11-, 10-H) ppm.

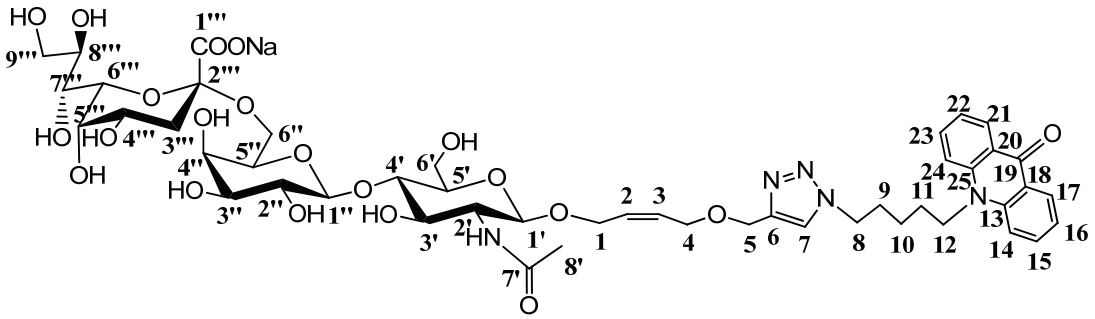
$^{13}$C NMR (126 MHz, Deuterium Oxide) δ=178.12 (C-19), 174.21 (C-1’”), 173.70 (C-7’”), 143.87 (C-6), 140.65 (C-13, -25), 134.57 (C-15, -23), 129.54 (C-2), 128.87 (C-3), 126.25 (C-17, -21), 124.66 (C-7), 121.66 (C-16, -21), 120.43 (C-18, -20), 115.05 (C-14, -24), 103.46 (C-1”), 100.12 (C-1’), 99.70 (C-2’”), 80.43 (C-4’), 74.63 (C-3”), 73.86 (C-8’”), 73.59 (C-5’), 72.53 (C-5”), 72.49 (C-3’), 72.28 (C-2’), 70.76 (C-6’”), 70.27 (C-7’”), 70.08 (C-5’”), 68.56 (C-4’”), 68.27 (C-4”), 65.15 (C-4), 64.54 (C-1), 63.65 (C-6”), 62.82 (C-9’”), 62.32 (C-5), 60.32 (C-6’), 54.93 (C-2”), 50.07 (C-8), 45.24 (C-12), 39.85 (C-3’”), 28.84 (C-9), 25.79 (C-11), 22.67 (C-10), 22.44 (C-8’) ppm.

**10-[5-(4-(((Z)-4-*O*-((3-deoxy-α-D- *glycero*-D-*gulo*-2-nonulopyranosonic acid)-(2→6)-β-D-galactopyranosyl-(1,4)-2-acetamido-2-deoxy-β-D-glucopyranosyloxy)-but-2-enyloxy)methyl)-1*H*-1,2,3-triazol-1-yl)pentyl]-10*H*-acridin-9-on**

**(*epi*-Kdn-2,6-LacNAc-D-T-P-Acr)**

For the structure identification of the product catalyzed by recombinant 2,6SiaT$_{ple}$, 25.2 mg

CTP (2 eq.) and 11 mg *epi*-Kdn (1.5 eq.) were added into 50 mL Tris-HCl (pH8.6) buffer with 0.5 M NaCl and 0.1% Triton ×100. 5 mg CSS and 20 $\mu$l PPase were added to start the reaction. When the CMP-activation was completed, 20 mg LacNAc-D-T-P-Acr (1 eq.) and 5 mg recombinant 2,6Sia$T_{ple}$ were added to start the reaction. The pH was being adjusted for keeping at 8.6 by pH-stat during the reaction for 2-3 h. When the reaction finished, 300 mL -20°C methanol was added to stop the reaction and precipitate the most of proteins in the reaction system. The precipitate was removed by filtration. The filtrate was concentrated to 10 mL and loaded on Biogel P-2 (Bio-Rad, Germany) column (3 cm × 100 cm). The products were further washed out by $ddH_2O$ and collected and lyophilized for NMR determination. Product: 23.1 mg yellow powder. Yield: 86%.



Chemical Formula: $C_{48}H_{64}N_5NaO_{21}$
Molecular Weight: 1070.03

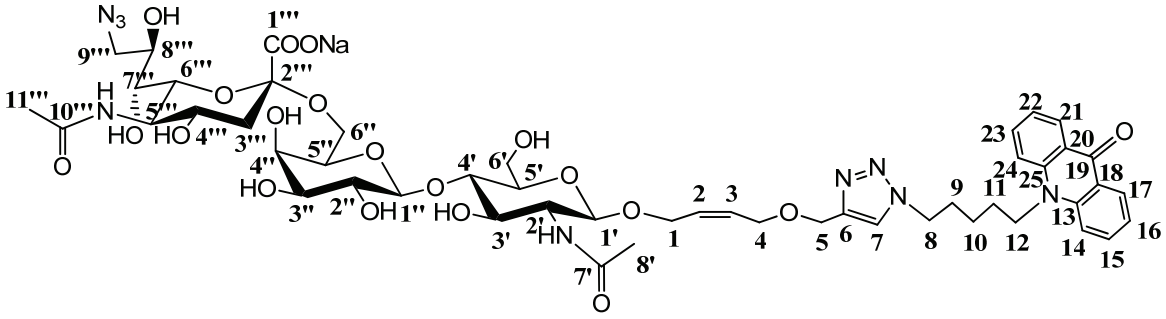
$^1$H NMR (500 MHz, Deuterium Oxide) δ=7.84 (dd, *J* = 8.1, 1.7 Hz, 2H, 17- 21H), 7.76 (s, 1H, 7-H), 7.34 (ddd, *J* = 8.5, 6.7, 1.6 Hz, 2H, 15-, 23-H), 6.99 – 6.86 (m, 4H, 14-, 24-, 16-, 22-H), 5.62 (q, *J* = 5.2 Hz, 2H, 2-, 3-H), 4.47 (s, 2H, 5-H), 4.39 (d, *J* = 7.9 Hz, 1H, 1”-H), 4.26 – 4.06 (m, 4H, 1’, 8-, 1a-H), 4.05 – 3.72 (m, 12H, 1b-, 4-, 9’”a-, 5’”-, 8’”-, 6’”-, 6’a-, 6”a-, 4”-, 4’”-, 7’”-H), 3.72 – 3.48 (m, 10H, 5”-, 6’b-, 6”b-, 3”-, 9’”-, 12-, 3’-, 4’-, 2”-H), 3.47 – 3.42 (m, 1H, 5’-H), 2.39 (dd, *J* = 12.2, 4.6 Hz, 1H, 3’”a-H), 1.97 (s, 3H, 8’-H), 1.85 (t, *J* = 12.4 Hz, 1H, 3’”-H), 1.65 – 1.51 (m, 2H, 9-H), 1.16 – 0.95 (m, 4H, 11-, 10-H) ppm.

$^{13}$C NMR (126 MHz, Deuterium Oxide) δ=178.08 (C-19), 143.86 (C-6), 140.63 (C-13, -25), 134.55 (C-15, -23), 129.52 (C-2), 128.86 (C-3), 126.23 (C-17, -21), 124.65 (C-7), 121.64 (C-16, -21), 120.42 (C-18, -20), 115.02 (C-14, -24), 103.48 (C-1”), 100.62 (C-1’), 99.83 (C-2’”), 80.26 (C-4’), 74.66 (C-3”), 73.77 (C-8’”), 73.17 (C-5’), 72.59 (C-5”), 72.50 (C-3’), 71.99 (C-2’), 71.12 (C-6’”), 70.84 (C-7’”), 69.46 (C-5’”), 68.51 (C-4’”), 67.61 (C-4”), 65.16 (C-4), 64.55 (C-1), 63.22 (C-6”), 62.31 (C-9’”), 62.21 (C-5), 60.32 (C-6’), 54.88 (C-2”), 50.06 (C-8), 45.23 (C-12), 34.54 (C-3’”), 28.84 (C-9), 25.78 (C-11), 22.66 (C-10), 22.49 (C-8’) ppm.

**10-[5-(4-(((Z)-4-*O*-(( 9-aido-*N*-acetyl-α-D-neuraminic acid)-(2→6)-β-D- galactopyranosyl-(1,4)-2-acetamido-2-deoxy-β-D-glucopyranosyloxy)-but-2-enyloxy)methyl)-1*H*-1,2,3-triazol-1-yl)pentyl]-10*H*-acridin-9-on**

**(Neu5Ac9$N_3$-2,6-LacNAc-D-T-P-Acr)**

For the structure identification of the product catalyzed by recombinant 2,6SiaT$_{ple}$, 25.2 mg CTP (2 eq.) and 13.5 mg Neu5Ac9$N_3$ (1.5 eq.) were added into 50 mL Tris-HCl (pH8.6) buffer with 0.5 M NaCl and 0.1% Triton ×100. 5 mg CSS and 20 μl PPase were added to start the reaction. When the CMP-activation was completed, 20 mg LacNAc-D-T-P-Acr (1 eq.) and 5 mg recombinant 2,6SiaT$_{ple}$ were added to start the reaction. The pH was being adjusted for keeping at 8.6 by pH-stat during the reaction for 2-3 h. When the reaction finished, 300 mL -20°C methanol was added to stop the reaction and precipitate the most of proteins in the reaction system. The precipitate was removed by filtration. The filtrate was concentrated to 10 mL and loaded on Biogel P-2 (Bio-Rad, Germany) column (3 cm × 100 cm). The products were further washed out by dd$H_2O$ and collected and lyophilized for NMR determination. Product: 22.2 mg yellow powder. Yield: 78%.



Chemical Formula: $C_{50}H_{66}N_9NaO_{20}$
Molecular Weight: 1136.10

$^1$H NMR (500 MHz, Deuterium Oxide) δ=7.81 – 7.71 (m, 2H, 17-, 21-H), 7.65 (s, 1H, 7-H), 7.33 – 7.15 (m, 2H, 15-, 23-H), 6.93 – 6.76 (m, 4H, 14-, 24-, 16-, 22-H), 5.58 – 5.43 (m, 2H, 2-, 3-H), 4.37 (s, 2H, 5-H), 4.34 (d, *J* = 8.1 Hz, 1H, 1”-H), 4.27 (d, *J* = 7.8 Hz, 1H, 1’-H), 4.15 – 4.03 (m, 3H, 8-, 1a-H), 4.00 (dd, *J* = 12.8, 5.8 Hz, 1H, 1b-H), 3.92 – 3.78 (m, 5H, 4-, 8”’-, 6”’-, 6”a-H), 3.78 – 3.60 (m, 5H, 5”’-, 5”-, 6’-, 6”a-, 4”-H), 3.60 – 3.37 (m, 11H, 4”’-, 6”b-, 3”-, 9”’a-, 12-, 7”’-, 3’-, 4’-, 2”-H), 3.36 – 3.29 (m, 2H, 2’-, 9”’b-H), 2.54 (dd, *J* = 12.3, 4.7 Hz,

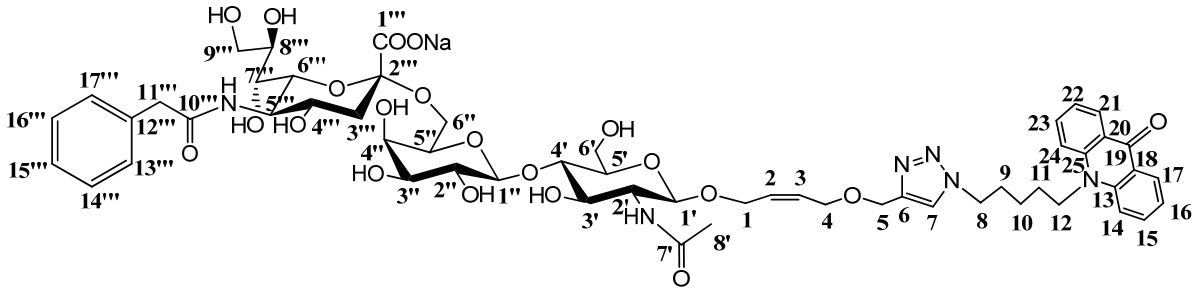
1H, 3'''a-H), 1.90 (s, 3H, 8'-H), 1.88 (s, 2H, 11'''-H), 1.57 (t, $J$ = 12.2 Hz, 1H, 3'''-H), 1.52-1.40 (m, 2H, 9-H), 1.11 – 0.80 (m, 4H, 11-, 10-H) ppm.

$^{13}$C NMR (126 MHz, Deuterium Oxide) δ=178.13 (C-19), 174.90 (C-1'''), 174.33 (C-7'''), 173.52 (C-10'''), 143.86 , 140.67 , 134.57 , 129.55 , 128.87 , 126.26 , 124.65 , 121.65 , 120.45 , 115.06 , 103.56 (C-1''), 100.29 (C-1'), 99.70 (C-2'''), 80.73 (C-4'), 74.58 (C-3''), 73.74 (C-8'''), 72,65 (C-5'), 72.55 (C-5''), 72.44 (C-3'), 70.78 (C-2'), 70.34 (C-6'''), 69.18 (C-7'''), 68.44 (C-4'''), 68.24 (C-4''), 65.14 (C-4), 64.53 (C-1), 63.39 (C-6''), 62.31 (C-5), 60.41 (C-6'), 54.84 (C-2''), 53.26 (C-9'''), 52.00 (C-5'''), 50.06 (C-8), 45.22 (C-12), 40.21 (C-3'''), 28.85 (C-9), 25.80 (C-11), 22.67 (C-10), 22.40 (C-8'), 22.14 (C-11''') ppm.

**10-[5-(4-(((Z)-4-*O*-((*N*-phenylacetyl-α-D-neuraminic acid)-(2→6)-β-D- galactopyranosyl-(1,4)-2-acetamido-2-deoxy-β-D-glucopyranosyloxy)-but-2-enyloxy)methyl)-1*H*-1,2,3-triazol-1-yl)pentyl]-10*H*-acridin-9-on**

**(Neu5NPhAc-2,6-LacNAc-D-T-P-Acr)**

For the structure identification of the product catalyzed by recombinant 2,6SiaT$_{ple}$, 25.2 mg CTP (2 eq.) and 15.3 mg Neu5NPhAc (1.5 eq.) were added into 50 mL Tris-HCl (pH8.6) buffer with 0.5 M NaCl and 0.1% Triton ×100. 5 mg CSS and 20 μl PPase were added to start the reaction. When the CMP-activation was completed, 20 mg LacNAc-D-T-P-Acr (1 eq.) and 5 mg recombinant 2,6SiaT$_{ple}$ were added to start the reaction. The pH was being adjusted for keeping at 8.6 by pH-stat during the reaction for 2-3 h. When the reaction finished, 300 mL -20°C methanol was added to stop the reaction and precipitate the most of proteins in the reaction system. The precipitate was removed by filtration. The filtrate was concentrated to 10 mL and loaded on Biogel P-2 (Bio-Rad, Germany) column (3 cm × 100 cm). The products were further washed out by dd$H_2O$ and collected and lyophilized for NMR determination. Product: 15.2 mg yellow powder. Yield: 51%.



Chemical Formula: $C_{56}H_{71}N_6NaO_{21}$
Molecular Weight: 1187.18

$^1$H NMR (500 MHz, Deuterium Oxide) δ=7.74 (dd, *J* = 8.1, 1.6 Hz, 2H, 17-, 21-H), 7.59 (s, 1H, 7-H), 7.25 – 7.13 (m, 2H, 13'"-, 17'"-H), 7.10 – 6.92 (m, 5H, 15'"-, 14'"-, 16'"-, 15-, 23-H), 6.86 – 6.70 (m, 4H, 24-, 14-, 16-, 22-H), 5.52 – 5.42 (m, 2H, 2-, 3-H), 4.31 (s, 2H, 5-H), 4.23 (d, *J* = 7.9 Hz, 1H, 1"-H), 4.04 (q, *J* = 5.2 Hz, 3H, 8-, 1a-H), 3.99 – 3.92 (m, 1H, 1b-H), 3.88 – 3.71 (m, 5H, 1b-, 4-, 9'"a-, 5'"-H), 3.71 – 3.46 (m, 10H, 8'"-, 6'"-, 6-, 6"a-, 4"-, 4'"-, 5"-, 11'"-H), 3.46 – 3.29 (m, 9H, 6"b-, 3"-, 9'"-, 12-, 7'"-, 3'-, 4'-, 2"-H), 3.16 (dd, *J* = 8.9, 1.7 Hz, 1H, 2'-H), 2.52 (dd, *J* = 12.4, 4.3 Hz, 1H, 3'"a-H), 1.73 (s, 3H, 8'-H), 1.56 (t, *J* = 11.8 Hz, 1H, 3'"b-H), 1.47 – 1.33 (m, 2H, 9-H), 1.05 – 0.80 (m, 4H, 11-, 10-H) ppm.

$^{13}$C NMR (126 MHz, Deuterium Oxide) δ=178.10 (C-19), 175.30 (C-1'"), 174.22 (C-7'"), 173.62 (C-10'"), 143.90 (C-6), 140.69 (C-13, -25), 135.06 (C-15, -23), 134.54 (C-12'"), 129.71 (C-15'"), 129.01 (C-2), 128.76 (C-3), 128.71 (C-14'", -16'"), 127.12 (C-13'", -17'"), 126.31 (C-17, -21), 124.59 (C-7), 121.63 (C-16, -22), 120.53 (C-18, -20), 115.03 (C-14, -24), 103.63 (C-1"), 100.07 (C-1'), 99.80 (C-2'"), 81.04 (C-4'), 74.53 (C-3"), 73.80 (C-8'"), 72.60 (C-5'), 72.57 (C-5"), 72.54 (C-3'), 71.86 (C-2'), 70.78 (C-6'"), 68.67 (C-7'"), 68.47 (C-4'"), 67.96 (C-4"), 65.27 (C-4), 64.64 (C-1), 63.40 (C-6"), 62.90 (C-9'"), 62.34 (C-5), 60.46 (C-6'), 54.81 (C-2"), 52.27 (C-5'"), 50.02 (C-8), 45.21 (C-12), 42.60 (C-11'"), 40.30 (C-3'"), 28.90 (C-9), 25.84 (C-11), 22.71 (C-10), 22.38 (C-8') ppm.

## 13.9. Cloning, Expression and Characterization of a Potential 2,3-Sialyltransferase from *Shewanella piezotolerans*

### 13.9.1. Cloning and expression of SiaTspi

#### 13.9.1.1. Cloning of $SiaT_{spi}$-U

The DNA sequence of SiaTspi gene from *Shewanella piezotolerans* with upstream sequence ($SiaT_{spi}$-U) was optimized for expression in *E. coli* and synthesized by Mr. Gene Company. *Nde*I and *Bam*HI restriction enzyme sites were added at 5'- and 3'-terminuses, respectively, for further cloning work. The plasmid pMA-SiaTspi-U containing synthesized $SiaT_{spi}$-U gene from Mr. Gene Company was transformed into *E. coli* Jm109 competent cells for plasmid amplification. The recombinant cells were cultured in 5 mL LB medium comprising 50 $\mu$g/mL ampicillin at 37°C with 250 rpm shaking overnight. The amplified plasmid was extracted and digested by *Nde*I and *Bam*HI, and then purified by gel extraction. After that, the $SiaT_{spi}$-U gene fragment was ligated into pET19b vector to construct the expression plasmid pET-SiaTspi-U, which was then transformed into *E. coli* Jm109 competent cells for plasmid amplification and identification. The positive colonies were picked and cultured in 5 mL Amp-LB medium at 37°C with 250 rpm shaking overnight. The extracted plasmids were identified by *Nde*I and *Bam*HI restriction endonuclease enzyme digestion and electrophoresis, as well as sequencing. Then, the identified final plasmid was transformed into *E. coli* BL21 Tuner competent cells. The positive colony (BL21T-pET-$SiaT_{spi}$-U) was picked and cultured in 5 mL Amp-LB medium at 37°C with 250 rpm shaking overnight and stored in 16% glycerol at -80°C.

```
1       CATATGGTTG CCAGCAGCAG CTCAAATGCC CAGCAGTCTC CGTATGCCCT
        GGAAATCTAT ATCGACAAAG CCACTCTGCC TTCAGTACAA CAAGCGGCCA
101     TGCTGGTGAA TAATCAGTCC CATAACCCGA AACTGATTTG TTGGCAACGT
        CATCCGGTGA ACGATGAAGC CCTGCTGCAG GGTATTAATG CCGCCTCTTT
201     CGTGTCTATT GCCTCACTGT GTCAGCACGC CGCCACTCTG CTGGCTGGTC
        ATCCTCATAG CCATATCACC ATTTATGGTA ACACCTATTG GTCTAAAGAT
301     CTGGCCCGTC TGATCCGTTA TCTGACTCGT ATCAGCGGTG TGGAGATCAA
        AAAACTGGAG CTGATCGACG ATGGTAGCAG CGAATATCAA AAAATGTTCT
```

```
401     ATTGGCAGCG CCTGTCTAGT GAAGAGCAAA CCCGTGACCT GGCCACTGGT
        CTGAAAAACC TGAAATCCTA TCTGTCCGGG AACGATAATA AACTGCTGCG
501     TCTGCTGACC GGACATTCGA ACAAACTGCC ACGTCGTCTG TCATCTTTTA
        TGAACTGGCA CCAACTGTTT CCGACGACCT ATCACATGCT GCGTATGGAC
601     TATCTGGATA AACCAGAACT GCATCAACTG AAACAATATC TGGGGAACAA
        TGCTCAACAA ATTCGCTGGA ACTATATTGC CGACAACCTG TTCGATGATG
701     AACAACAGTC CCTGTTCTAT CAACTGCTGG GAATCTCTCT GGCAGAACAA
        AAACAACTGC GTGCGGGACG TCAACAACTG CACGACTTCA TGTTTATCGG
801     GGTCGATTCA TCCAATGCCT CCAGCAAACT GCAGATTAAC GTGATTGCCG
        ATTCTCGTCA AGAAAGCGGA ATCATTCCTA CGATTACCGC TAAAAAAATG
901     CTGTTCAAAG GCCACCCTTT CGCCAATTTC AATCAAACGA TCGTGGATGC
        CCATCAAATG GGAGAAATGC CGGCCATGAT CCCTTTCGAA ACCCTGATCA
1001    TGACAGGAAA CCTGCCTCAG AAAGTCGGTG GAATGGCATC AAGTCTGTAT
        TTTTCCCTGC CGAACAACTA TCATATCGAG TATATCGTGT TCTCTGGCTC
1101    CAAAAAAGAC CTGGAGCAAC ATGCTCTGCT GCAAATCATG CTGTATCTGA
        AAGTGATTAG CCCGGAACGT GTGTATTTTA GTGAACAATT CAAATCCTGC
1201    TAAGGATCC
```

Sequence 13.5. Artificial gene sequence of SiaTspi-U

### 13.9.1.2. Cloning of $SiaT_{spi}$

The pET19b-$SiaT_{spi}$-U was prepared as template for the PCR of the gene modification. Primers d31SWSTf 5'-CAT ATG CTG GTG AAT AAT CAG TTC CAT-3' and d31SWSTb 5'-GGA TCC TTA GCA GGA TTT GAA TTG TTC-3' were designed synthesized by Biomers (Germany). The PCR reaction used Phusion High-Fidelity DNA Polymerase (Fermentas, Germany).

PCR reaction:

| | |
|---|---|
| dsDNA template: | 2 $\mu$l |
| Primer 1: | 2 $\mu$l |
| Primer 2: | 2 $\mu$l |
| dNTP mix: | 1 $\mu$l |
| 10×Reaction Buffer: | 5 $\mu$l |
| DNA Polymerase: | 0.5 $\mu$l |
| $ddH_2O$: | up to 50 $\mu$l |

| | | |
|---|---|---|
| 97°C | 2 min | |
| 97°C | 10 sec | |
| 68°C | 10 sec | |
| 72°C | 4 min | 30 cycles |
| 72°C | 10 min | |

The PCR products were identified by electrophoresis and purified by gel extraction. Then, they were ligated into pUC19 vector for sub cloning. The ligated products were transformed into Jm109 competent cells and cultured on Amp-LB-agar plates at 37°C overnight. The positive colonies were picked and cultured in 5 mL amp-LB medium at 37°C with 250 rpm shaking overnight. Then their plasmids were extracted and identified by *Nde*I and *Bam*HI restriction endonuclease enzyme digestion and electrophoresis. The positive plasmid was digested by *Nde*I and *Bam*HI at 37°C overnight. The SiaTspi gene fragments were purified by gel extraction and then ligated into pET19b vector to obtain pET19b-SiaTspi plasmid. The ligation products were transformed into Jm109 competent cells and cultured on Amp-LB-agar plates at 37°C overnight. The positive colonies were picked and cultured in 5 mL Amp-LB medium at 37°C with 250 rpm shaking overnight. Then their plasmids were extracted and identified by *Nde*I and *Bam*HI restriction endonuclease enzyme digestion and electrophoresis, as well as sequencing. The positive plasmid was then transformed into BL21 Tuner competent cells and cultured on Amp-LB-agar plates at 37°C overnight. The positive colonies were picked and cultured in 5 mL amp-LB medium at 37°C with 250 rpm shaking overnight, and then stored in 16% glycerol at -80°C.

```
1       ATGCTGGTGA ATAATCAGTC CCATAACCCG AAACTGATTT GTTGGCAACG
        TCATCCGGTG AACGATGAAG CCCTGCTGCA GGGTATTAAT GCCGCCTCTT
101     TCGTGTCTAT TGCCTCACTG TGTCAGCACG CCGCCACTCT GCTGGCTGGT
        CATCCTCATA GCCATATCAC CATTTATGGT AACACCTATT GGTCTAAAGA
201     TCTGGCCCGT CTGATCCGTT ATCTGACTCG TATCAGCGGT GTGGAGATCA
        AAAAACTGGA GCTGATCGAC GATGGTAGCA GCGAATATCA AAAAATGTTC
301     TATTGGCAGC GCCTGTCTAG TGAAGAGCAA ACCCGTGACC TGGCCACTGG
        TCTGAAAAAC CTGAAATCCT ATCTGTCCGG GAACGATAAT AAACTGCTGC
401     GTCTGCTGAC CGGACATTCG AACAAACTGC CACGTCGTCT GTCATCTTTT
        ATGAACTGGC ACCAACTGTT TCCGACGACC TATCACATGC TGCGTATGGA
501     CTATCTGGAT AAACCAGAAC TGCATCAACT GAAACAATAT CTGGGGAACA
        ATGCTCAACA AATTCGCTGG AACTATATTG CCGACAACCT GTTCGATGAT
```

```
601     GAACAACAGT CCCTGTTCTA TCAACTGCTG GGAATCTCTC TGGCAGAACA
        AAAACAACTG CGTGCGGGAC GTCAACAACT GCACGACTTC ATGTTTATCG
701     GGGTCGATTC ATCCAATGCC TCCAGCAAAC TGCAGATTAA CGTGATTGCC
        GATTCTCGTC AAGAAAGCGG AATCATTCCT ACGATTACCG CTAAAAAAAT
801     GCTGTTCAAA GGCCACCCTT TCGCCAATTT CAATCAAACG ATCGTGGATG
        CCCATCAAAT GGGAGAAATG CCGGCCATGA TCCCTTTCGA AACCCTGATC
901     ATGACAGGAA ACCTGCCTCA GAAAGTCGGT GGAATGGCAT CAAGTCTGTA
        TTTTTCCCTG CCGAACAACT ATCATATCGA GTATATCGTG TTCTCTGGCT
1001    CCAAAAAAGA CCTGGAGCAA CATGCTCTGC TGCAAATCAT GCTGTATCTG
        AAAGTGATTA GCCCGGAACG TGTGTATTTT AGTGAACAAT TCAAATCCTG
1101    CTAA
```

Sequence 13.6. Artificial gene sequence of $SiaT_{spi}$

### 13.9.1.3. Cloning of $dSiaT_{spi}$-U

The pET19b-SiaTspi was prepared as template for the PCR of the gene modification. Primers dSMSTUf 5'-TCG AAC AAA CTG CCA CGT CGT-3' and dSMSTUb 5'-CCC GGA CAG ATA GGA TTT CAG GT-3' were designed synthesized by Biomers (Germany). The PCR reaction used Phusion High-Fidelity DNA Polymerase (Fermentas, Germany).

PCR reaction:

| | |
|---|---|
| dsDNA template: | 2 $\mu$l |
| Primer 1: | 2 $\mu$l |
| Primer 2: | 2 $\mu$l |
| dNTP mix: | 1 $\mu$l |
| 10×Reaction Buffer: | 5 $\mu$l |
| DNA Polymerase: | 0.5 $\mu$l |
| $ddH_2O$: | up to 50 $\mu$l |

| | | |
|---|---|---|
| 97°C | 2 min | |
| 97°C | 10 sec | |
| 68°C | 10 sec | |
| 72°C | 4 min | 30 cycles |
| 72°C | 10 min | |

The PCR products were identified by electrophoresis and purified by gel extraction. Then, they were ligated into pUC19 vector for sub cloning. The ligated products were transformed into Jm109 competent cells and cultured on Amp-LB-agar plates at 37°C overnight. The positive colonies were picked and cultured in 5 mL amp-LB medium at 37°C with 250 rpm shaking overnight. Then their plasmids were extracted and identified by *Nde*I and *Bam*HI restriction endonuclease enzyme digestion and electrophoresis. The positive plasmid was digested by *Nde*I and *Bam*HI at 37°C overnight. The dSiaTspi-U gene fragments were purified by gel extraction and then ligated into pET19b vector to obtain pET19b-dSiaT$_{spi}$-U plasmid, respectively. The ligation products were transformed into Jm109 competent cells and cultured on Amp-LB-agar plates at 37°C overnight. The positive colonies were picked and cultured in 5 mL Amp-LB medium at 37°C with 250 rpm shaking overnight. Then their plasmids were extracted and identified by *Nde*I and *Bam*HI restriction endonuclease enzyme digestion and electrophoresis, as well as sequencing. The positive plasmid was then transformed into BL21 Tuner competent cells and cultured on Amp-LB-agar plates at 37°C overnight. The positive colonies were picked and cultured in 5 mL amp-LB medium at 37°C with 250 rpm shaking overnight, and then stored in 16% glycerol at -80°C.

```
1       CATATGGTTG CCAGCAGCAG CTCAAATGCC CAGCAGTCTC CGTATGCCCT
        GGAAATCTAT ATCGACAAAG CCACTCTGCC TTCAGTACAA CAAGCGGCCA
101     TGCTGGTGAA TAATCAGTCC CATAACCCGA AACTGATTTG TTGGCAACGT
        CATCCGGTGA ACGATGAAGC CCTGCTGCAG GGTATTAATG CCGCCTCTTT
201     CGTGTCTATT GCCTCACTGT GTCAGCACGC CGCCACTCTG CTGGCTGGTC
        ATCCTCATAG CCATATCACC ATTTATGGTA ACACCTATTG GTCTAAAGAT
301     CTGGCCCGTC TGATCCGTTA TCTGACTCGT ATCAGCGGTG TGGAGATCAA
        AAAACTGGAG CTGATCGACG ATGGTAGCAG CGAATATCAA AAAATGTTCT
401     ATTGGCAGCG CCTGTCTAGT GAAGAGCAAA CCCGTGACCT GGCCACTGGT
        CTGAAAAACC TGAAATCCTA TCTGTCCGGG TCGAACAAAC TGCCACGTCG
501     TCTGTCATCT TTTATGAACT GGCACCAACT GTTTCCGACG ACCTATCACA
        TGCTGCGTAT GGACTATCTG GATAAACCAG AACTGCATCA ACTGAAACAA
601     TATCTGGGGA ACAATGCTCA ACAAATTCGC TGGAACTATA TTGCCGACAA
        CCTGTTCGAT GATGAACAAC AGTCCCTGTT CTATCAACTG CTGGGAATCT
701     CTCTGGCAGA ACAAAAACAA CTGCGTGCGG GACGTCAACA ACTGCACGAC
        TTCATGTTTA TCGGGGTCGA TTCATCCAAT GCCTCCAGCA AACTGCAGAT
801     TAACGTGATT GCCGATTCTC GTCAAGAAAG CGGAATCATT CCTACGATTA
        CCGCTAAAAA AATGCTGTTC AAAGGCCACC CTTTCGCCAA TTTCAATCAA
```

```
901     ACGATCGTGG ATGCCCATCA AATGGGAGAA ATGCCGGCCA TGATCCCTTT
        CGAAACCCTG ATCATGACAG GAAACCTGCC TCAGAAAGTC GGTGGAATGG
1001    CATCAAGTCT GTATTTTTCC CTGCCGAACA ACTATCATAT CGAGTATATC
        GTGTTCTCTG GCTCCAAAAA AGACCTGGAG CAACATGCTC TGCTGCAAAT
1101    CATGCTGTAT CTGAAAGTGA TTAGCCCGGA ACGTGTGTAT TTTAGTGAAC
        AATTCAAATC CTGCTAAGGA TCC
```

Sequence 13.7. Artificial gene sequence of dSiaTspi-U

#### 13.9.1.4. Expression of $SiaT_{spi}$, $SiaT_{spi}$-U and $dSiaT_{spi}$-U

10 $\mu$l BL21T-pET21a-$SiaT_{spi}$, BL21T-pET19b-$dSiaT_{spi}$-U and BL21T-pET19b-$dSiaT_{spi}$-U were transferred into 5 mL LB medium, respectively, and cultured at 37°C with 250 rpm shaking until OD reached 0.5. 0.1 mM IPTG was added into the culture for the expression induction at 28°C with shaking overnight. After that, the cells were harvested by centrifugation at 4,000 rpm for 20 min. The expression levels of the recombinant $SiaT_{spi}$, $SiaT_{spi}$-U and $dSiaT_{spi}$-U were identified by SDS-PAGE. The identification of recombinant $SiaT_{spi}$, $SiaT_{spi}$-U and $dSiaT_{spi}$-U activities were used the method described in 13.5.1.3.

### 13.9.2.Optimization of expression of $SiaT_{spi}$ in *E. coli*

#### 13.9.2.1. Coexpression with chemprons

The chaperone plasmids pG-KJE8, pGro7, pKJE7, pG-Tf2 and pTf16 were transformed into BL21DE3 competent cells and cultured on LB-agar plates with chloramphenicol. The positive colonies were picked and cultured in LB medium with chloramphenicol at 37ºC overnight. Then the recombinants were stored in 16% glycerol at -80ºC.

Chaperone BL21(DE3) cells was cultured in 5 ml LB medium with chloramphenicol at 37ºC with 250 rpm shaking for 16-18 h. 2.5 ml each culture was transferred into 250 ml SOB medium and cultured at 37ºC with 250 rpm shaking until $OD_{600}$ reaches 0.6-0.8. Then the cultures were put in ice for 10 min. After that, the cells were harvested by 2,000g

centrifugation at 4°C for 10 min, followed by adding 80 ml cold TB medium to resuspend the cells. The cells were put in ice for 10 min again and centrifuged the resuspension at 2,000g 4°C for 10 min to remove the TB medium. Then 20ml TB medium was added and the cells were mixed gently. The resuspension was treated by ice for 10 min. Finally, 1.5 ml DMSO and 75 mM DTT were added and mixed gently. The cells were frozen by liquid nitrogen and stored at -80°C.

The pET21a-SiaT$_{spi}$, pET19b-SiaT$_{spi}$-U and pET19b-dSiaT$_{spi}$-U plasmids were transformed into the competent chaperone cells and cultured on LB agar plate with ampicillin and chloramphenicol at 37°C overnight. The positive colonies were picked and identified by plasmid identification. After that, these five recombinant cells were stored in 16% glycerol and stored at -80°C.

10 $\mu$l Recombinant SiaT$_{spi}$-chaperone BL21(DE3) cells were transferred into 5 ml LB-amp medium and cultured at 37°C 250 rpm until OD$_{600}$ reached 0.5. 0.1 mM IPTG was added to induce the recombinant protein expression. The induction was taken place at 28°C 250 rpm overnight. The cells were harvested by centrifuge and lysate by ultrasonic treatment. The soluble expression level was identified by SDA-PAGE.

#### 13.9.2.2. Characterization of SiaT$_{spi}$

The identification of recombinant SiaT$_{spi}$, SiaT$_{spi}$-U and dSiaT$_{spi}$-U activities were used the method described in 13.5.1.3.

#### 13.9.2.3. Optimization of expression at low temperature

10 $\mu$l Recombinant SiaTspi-chaperone BL21(DE3) cells were transferred into 5 ml LB-amp medium and cultured at 37°C 250 rpm until OD600 reached 0.5. 0.1 mM IPTG was added to induce the recombinant protein expression. The induction was taken place at 16°C 250 rpm overnight. The cells were harvested by centrifuge and lysate by ultrasonic treatment. The soluble expression level was identified by SDS-PAGE.

### 13.9.2.4. Expression of $SiaT_{spi}$ in *Yarrowia lipolytica* expression system

*a) Expression plasmid contruction*

p64PT vector was transformed into *E. coli* Jm109 competent cells for plasmid amplification. The recombinant cells were cultured in 5 mL LB medium comprising 50 $\mu$g/mL ampicillin at 37°C with 250 rpm shaking overnight. The plasmid was extracted and digested by *Spn*I and then purified by gel extraction.

The pET19b-SiaTspi-U and pET19b-dSiaTspi were prepared as templates for the PCR of the SiaTspi-U and dSiaTspi-U amplification. Primers p64SiaTspi-f 5'- ATA TGC ATG CTC CAT CAT CAT CAT CAT CAT C-3' and p64SiaTspi-b 5'- ATA TGC ATG CTT AGC AGG ATT TGA ATT GTT CAC-3' were designed synthesized by Biomers (Germany). The PCR reaction used Phusion High-Fidelity DNA Polymerase (Fermentas, Germany).

PCR reaction:

| | |
|---|---|
| dsDNA template: | 2 $\mu$l |
| Primer 1: | 2 $\mu$l |
| Primer 2: | 2 $\mu$l |
| dNTP mix: | 1 $\mu$l |
| 10×Reaction Buffer: | 5 $\mu$l |
| DNA Polymerase: | 0.5 $\mu$l |
| $ddH_2O$: | up to 50 $\mu$l |

| | | |
|---|---|---|
| 97°C | 2 min | |
| 97°C | 10 sec | |
| 65°C | 10 sec | |
| 72°C | 4 min | 30 cycles |
| 72°C | 10 min | |

The PCR products were identified by electrophoresis and purified by gel extraction. Then, they were ligated into pUC19 vector for sub cloning. The ligated products were transformed into Jm109 competent cells and cultured on Amp-LB-agar plates at 37°C overnight. The positive colonies were picked and cultured in 5 mL amp-LB medium at 37°C with 250 rpm shaking overnight. Then their plasmids were extracted and identified by *Spn*I restriction endonuclease enzyme digestion and electrophoresis. The positive plasmid was digested by

*Spn*I at 37°C overnight. The $SiaT_{spi}$-U and $dSiaT_{spi}$-U gene fragments were purified by gel extraction and then ligated into p64PT vector to obtain p64SiaTspi-U and p64dSiaTspi-U plasmids, respectively. The ligation products were transformed into Jm109 competent cells and cultured on Amp-LB-agar plates at 37°C overnight. The positive colonies were picked and cultured in 5 mL Amp-LB medium at 37°C with 250 rpm shaking overnight. Then their plasmids were extracted and identified by *Spn*I and *Nde*I restriction endonuclease enzyme digestion and electrophoresis independently, as well as sequencing. The positive recombinants were stored in 16% glycerol at -80°C.

The plasmid p64SiaT-U and p64dSiaT-U were extracted from 50 ml culture. The concentration and purity of the plasmids were determined by spectrophotometry. The plasmids were diluted into 0.1 $\mu g/\mu l$ and linearized by *Sac*II overnight.

### *b) Transformation of Yarrowia*

*Yarrowia* cells were transferred into 5 ml YPD medium pH 4.0 and shake at 28°C 220 rpm for about 10-12 h. 50-100 $\mu$l cultures were transferred into 10 ml YPD medium pH 4.0 with cell density 1*105 /ml and incubated at 28°C 220 rpm overnight until the cell density reached 9*107 to 1.5*108 /ml. The cells were harvested by centrifuge 3500 rpm at 28°C for 5 min. The cells were washed by 10 ml TE buffer 2 times. After that, about 20 ml 100 mM LiNAc pH 6.0 was added to make the final cell density to $5\times10^7$ /ml. The suspension was incubated at 28°C for 1h. Then LiNAc was removed by centrifuge and 1/10V LiNAc was added to suspend the cells.

1 $\mu$g Plasmid was added into a 2 ml tube and heated at 60°C for 5 min and then put into ice. 5 $\mu$gl Denatured Carrier-DNA (50 $\mu$g) was added and mixed. 100 $\mu$l Competent *Yarrowia* H222-S4 and H355 cells was then added into tube and mixed gently. The tubes were incubated at 28°C for 15 min. Then 700 $\mu$l 40% PEG4000 was added and mixed gently. The tubes were incubated at 28°C with shaking for 1h. After that, cells were incubated at 39°C for 10 min followed by adding 1.2 ml 100 mM LiNAc pH6.0. 200 $\mu$l suspension was spread on plates for 2 weeks culture at 28°C with enough $O_2$.

*c) Isolation of positive recombinant Yarrowia*

Four colonies for each recombinant (H222-S4-SiaT-U, H222-S4-dSiaT-U, H355-SiaT-U and H355-dSiaT-U) were picked and streaked on plates for isolation. Single colonies for each recombinant were picked and transferred into 10 ml YPD medium for 2 days culture at 28°C. 800 $\mu$l of each culture was stored as glycerol stock with 20% glycerol. The rest cultures were used for screening and copy number measurement by PCR.

*d) Screening and copy number measurement by PCR*

10 ml 2 days culture was transferred into a 50 ml centrifugal tube followed by centrifugation at 3,500 for 5 min to remove the medium. The pellet was washed by 1 ml $H_2O$ and suspended in 500 $\mu$l lysate buffer in a 2 ml tube, which contains 100 mM Tris-HCl (pH 8.0), 50 mM EDTA and 1% w/v SDS. 400 $\mu$l Glass beams were added. The tubes were treated by FastPrep for 20 s 4 times. The lysate was transferred into a new tube and incubated at 65°C for 5 min and then on ice for 5 min. After that, 500 $\mu$l chloroform was added and centrifuged at 13,000 rpm for 2 min. The supernatant was transferred into a new tube, and 1 ml ioropropanol was added. The mixture was incubated at room temperature for 5 min, and then centrifuged at 13,000 rpm for 5 min. The pellet was washed by 200 $\mu$l cold ethanol and dissolved in 50 $\mu$l dd$H_2O$ with 1 $\mu$l 10 mg/ml RNase A and followed by incubation at 37°C for 2 h.

The genome DNA prepared above was determined the concentration. 100 ng genome DNA for each sample was used as template for identification. The PCR primers were p64SiaTspi-f 5'-ATA TGC ATG CTC CAT CAT CAT CAT CAT CAT C-3' and p64SiaTspi-b 5'- ATA TGC ATG CTT AGC AGG ATT TGA ATT GTT CAC-3'. The PCR reaction used Phusion High-Fidelity DNA Polymerase (Fermentas, Germany).

PCR reaction:

| | |
|---|---|
| dsDNA template: | 2 $\mu$l |
| Primer 1: | 2 $\mu$l |
| Primer 2: | 2 $\mu$l |
| dNTP mix: | 1 $\mu$l |
| 10×Reaction Buffer: | 5 $\mu$l |

| | |
|---|---|
| DNA Polymerase: | 0.5 $\mu$l |
| $ddH_2O$: | up to 50 $\mu$l |

| | | |
|---|---|---|
| 97°C | 2 min | |
| 97°C | 10 sec | |
| 72°C | 10 sec | |
| 72°C | 4 min | 30 cycles |
| 72°C | 10 min | |

*e) Fermantation of recobinant Yarrowia*

Recombinant *Y. li* was transferred into 50 ml YPD medium for pre-culture at 28°C 220 rpm for 12-24 h. The cells were harvested by centrifuge at 28°C 3500 rpm for 5min, and then washed by 10 ml 1×Reader buffer one time. The washed cells were resuspended by 10 ml 1×Reader buffer and measured the cell density at $OD_{600}$. The cells with final cell density of $OD_{600}$ = 1 were transferred into 100 ml fermentation medium (1×Reader buffer, 2% glucose, 100 mM potassium phosphate buffer pH 5.8, Iron solution, vitamin solution and $FeCl_3$) and cultured at 28°C 220 rpm. The pH of the culture was measured and adjusted to 4.0 every 3 h. When the glucose was used out, add 1% ethanol for induction. The cell density and pH of the culture were measured every 3 h. Cell samples in each 3 h were taken for expression and activity determination. The induction can continue for 12-15 h up to 30 h.

*f) Identification of acctivity*

The cells were harvested by centrifuge at 28°C 3500 rpm for 5min, and then washed by 10 ml TE buffer one time. After that, the cells were collected in 2 ml tubes with 1.5 ml glass beans. The tubes were fixed on vortex and shaked at maximum speed for 5 min. The crude extract was pipetted out into new tubes and centrifuged at 12,000 rpm for 10 min. The supernatant was taken out for activity determination.The rapid activity determination was destribed in 13.9.2.2.

## References


[1] Soetaert, W. and Vandamme, E. The impact of industrial biotechnology. *Biotech. J.* 2006; **1**:756-769.

[2] Lorenz, P. and Zinke, H. White biotechnology: differences in US and EU approaches? *Trends in Biotechnol.* 2005; **23**:570-574.

[3] Sheldon, R.A. Fundamentals of green chemistry: efficiency in reaction design. *Chem. Soc. Rev.* 2012; **41**:1437-1451.

[4] Wenda, S., Illner, S., Mell, A. and Kragl, U. Industrial biotechnology-the future of green chemistry? *Green Chem.* 2011; **13**:3007-3047.

[5] Drauz, K. Enzyme Catalysis in Organic Synthesis: A Comprehensive Handbook in 3 Vol. set. Wiley-VCH, Weinheim, 2012.

[6] Schedel, M. White biotechnology at Bayer HealthCare Product Supply: More than thirty years experience. *Chem-Ing-Tech.* 2006; **78**:485-489.

[7] Montero, J.M.S. White biotechnology and pharmaceutical industry. *An. Real. Acad. Nac. F.* 2007; **73**:501-535.

[8] Watson, W.J.W. How do the fine chemical, pharmaceutical, and related industries approach green chemistry and sustainability? *Green Chem.* 2012; **14**:251-259.

[9] Schoemaker, H.E., Mink, D. and Wubbolts, M.G. Dispelling the myths - Biocatalysis in industrial synthesis. *Science* 2003; **299**:1694-1697.

[10] Aime, C. and Coradin, T. Nanocomposites from biopolymer hydrogels: Blueprints for white biotechnology and green materials chemistry. *J. Polym. Sci. Pol. Phys.* 2012; **50**:669-680.

[11] Paul, N. Subject agency of renewable raw materials: Federal constitution advances White Biotechnology. *Chem-Ing-Tech.* 2006; **78**:181-182.

[12] Philippe, M., Didillon, B. and Gilbert, L. Industrial commitment to green and sustainable chemistry: using renewable materials & developing eco-friendly processes and ingredients in cosmetics. *Green Chem.* 2012; **14**:952-956.

[13] Sell, D., Puls, J. and Ulber, R. White biotechnology - Energy solutions for the future? *Chem. Unserer Zeit* 2007; **41**:108-116.

[14] Buchholz, K., Kasche, V. and Bornscheuer, U.T. *Biocatalysts and enzyme technology*, Wiley-VCH, Weinheim, 2005.

[15] Akkoyun, T., Arslan, A., Turhan, I. and Karhan, M. Immobilization techniques in biotechnology. *Curr. Opin. Biotech.* 2011; **22**:S61-S61.

[16] Wackett, L.P. Immobilization of microbial enzymes An annotated selection of World Wide Web sites relevant to the topics in Microbial Biotechnology. *Microb. Biotechnol.* 2010; **3**:729-730.

[17] Schmid, A., Dordick, J.S., Hauer, B., Kiener, A., Wubbolts, M. and Witholt, B. Industrial biocatalysis today and tomorrow. *Nature* 2001; **409**:258-268.

[18] Tao, F., Liu, Y.H., Luo, Q., Su, F., Xu, Y.Q., Li, F.L., Yu, B., Ma, C.Q. and Xu, P. Novel organic solvent-responsive expression vectors for biocatalysis: Application for development of an organic solvent-tolerant biodesulfurizing strain. *Bioresource Technol.* 2011; **102**:9380-9387.

[19] Tang, X.Y., Sun, H.L. and He, B.F. Organic Solvent Tolerant Bacteria and Enzymes for Application in Biocatalysis. *Prog. Chem.* 2009; **21**:2726-2733.

[20] Carrea, G. and Riva, S. *Organic synthesis with enzymes in non-aqueous media*, Wiley-VCH, Weinheim, 2008.

[21] Boy, M. and Voss, H. Reaction engineering of biocatalysis in microstructured multiphase systems. *Chem-Ing-Tech.* 1996; **68**:831-835.

[22] Karande, R., Schmid, A. and Buehler, K. Miniaturizing Biocatalysis: Enzyme-Catalyzed Reactions in an Aqueous/Organic Segmented Flow Capillary Microreactor. *Adv. Synth. Catal.* 2011; **353**:2511-2521.

[23] Yoon, S.K., Choban, E.R., Kane, C., Tzedakis, T. and Kenis, P.J.A. Laminar flow-based electrochemical microreactor for efficient regeneration of nicotinamide cofactors for biocatalysis. *J. Am. Chem. Soc.* 2005; **127**:10466-10467.

[24] Bornscheuer, U.T., Huisman, G.W., Kazlauskas, R.J., Lutz, S., Moore, J.C. and Robins, K. Engineering the third wave of biocatalysis. *Nature* 2012; **485**:185-194.

[25] Lorenz, P. Metagenomics for white biotechnology. *Chem. Ing. Tech.* 2006; **78**:461-468.

[26] Lutz, S. and Bornscheuer, U.T. *Protein engineering handbook*, Wiley-VCH, Weinheim,, 2009.

[27] Strohmeier, G.A., Pichler, H., May, O. and Gruber-Khadjawi, M. Application of Designed Enzymes in Organic Synthesis. *Chem. Rev.* 2011; **111**:4141-4164.

[28] Liese, A., Seelbach, K. and Wandrey, C. *Industrial biotransformations*, Wiley-VCH, Weinheim, 2006.

[29] Patel, R.N. *Stereoselective biocatalysis*, Marcel Dekker Inc, Basel, 2000.

[30] Gotor, V., Alfonso, I. and García-Urdiales, E. *Asymmetric organic synthesis with enzymes*, Wiley-VCH, Weinheim, 2008.

[31] Schenk, G., Duggleby, R.G. and Nixon, P.F. Properties and functions of the thiamin diphosphate dependent enzyme transketolase. *Int. J. Biochem. Cell B.* 1998; **30**:1297-1318.

[32] de la Haba, G., Leder, I.G. and Racker, E. Crystalline transketolase from baker's yeast: ioslation and properties. *J. Biol. Chem.* 1955; **214**:409-426.

[33] Lairson, L.L., Henrissat, B., Davies, G.J. and Withers, S.G. Glycosyltransferases: structures, functions, and mechanisms. *Annu. Rev. Biochem.* 2008; **77**:521-555.

[34] Rosenberg, A. *Biology of the sialic acids*, Springer, Berlin, 1995.

[35] Varki, A. *Essentials of glycobiology*, Cold Spring Harbor Laboratory Press, New York, 1999.

[36] Enders, D., Niemeier, O. and Henseler, A. Organocatalysis by *N*-heterocyclic, carbenes. *Chem. Rev.* 2007; **107**:5606-5655.

[37] Zeitler, K. Extending mechanistic routes in heterazolium catalysis-promising concepts for versatile synthetic methods. *Angew. Chem. Int. Edit.* 2005; **44**:7506-7510.

[38] Widmann, M., Radloff, R. and Pleiss, J. The Thiamine diphosphate dependent Enzyme Engineering Database: A tool for the systematic analysis of sequence and structure relations. *Bmc. Biochem.* 2010; **11**.

[39] Sprenger, G.A. and Pohl, M. Synthetic potential of thiamin diphosphate-dependent enzymes. *J. Mol. Catal. B-Enzym.* 1999; **6**:145-159.

[40] Schörken, U. and Sprenger, G.A. Thiamin-dependent enzymes as catalysts in chemoenzymatic syntheses. *BBA-Protein Struct. M.* 1998; **1385**:229-243.

[41] Iding, H., Siegert, P., Mesch, K. and Pohl, M. Application of α-keto acid decarboxylases in biotransformations. *BBA-Protein Struct. M.* 1998; **1385**:307-322.

[42] Schenk, G., Duggleby, R.G. and Nixon, P.F. Properties and functions of the thiamin diphosphate dependent enzyme transketolase. *Int. J. Biochem. Cell. Biol.* 1998; **30**:1297-1318.

[43] Wohlgemuth, R. C2-Ketol elongation by transketolase-catalyzed asymmetric synthesis. *J. Mol. Catal. B-Enzym.* 2009; **61**:23-29.

[44] Schenk, G., Layfield, R., Candy, J.M., Duggleby, R.G. and Nixon, P.F. Molecular evolutionary analysis of the thiamine-diphosphate-dependent enzyme, transketolase. *J. Mol. Evol.* 1997; **44**:552-572.

[45] Sprenger, G.A., Schorken, U., Sprenger, G. and Sahm, H. Transketolase A of *Escherichia coli* K12. Purification and properties of the enzyme from recombinant strains. *Eur. J. Biochem.* 1995; **230**:525-532.

[46] Bolte, J., Demuynck, C. and Samaki, H. Utilization of enzymes in organic chemistry: Transketolase catalyzed synthesis of ketoses. *Tetrahedron Lett.*1987; **28**:5525-5528.

[47] Hobbs, G.R., Lilly, M.D., Turner, N.J., Ward, J.M., Willets, A.J. and Woodley, J.M. Enzyme-catalysed carbon-carbon bond formation: use of transketolase from *Escherichia coli*. *J. Chem. Soc. Perk. T. 1* 1993; 165-166.

[48] Demuynck, C., Bolte, J., Hecquet, L. and Dalmas, V. Enzyme-Catalyzed Synthesis of Carbohydrates - Synthetic Potential of Transketolase. *Tetrahedron Lett.* 1991; **32**:5085-5088.

[49] Hecquet, L., Demuynck, C., Schneider, G. and Bolte, J. Enzymatic syntheses of ketoses: study and modification of the substrate specificity of the transketolase from Saccharomyces cerevisiae. *J. Mol. Catal. B-Enzym.* 2001; **11**:771-776.

[50] Littlechild, J., Turner, N., Hobbs, G., Lilly, M., Rawas, A. and Watson. H. Crystallization and preliminary X-ray crystallographic data with *Escherichia coli* transketolase. *Acta. Crystallographica D* 1995; **51**:1074-1076.

[51] Asztalos, P., Parthier, C., Golbik, R., Kleinschmidt, M., Hübner, G., Weiss, M.S., Friedemann, R., Wille, G. and Tittmann, K. Strain and Near Attack Conformers in Enzymic Thiamin Catalysis: X-ray Crystallographic Snapshots of Bacterial Transketolase in Covalent Complex with Donor Ketoses Xylulose 5-phosphate and Fructose 6-phosphate, and in Noncovalent Complex with Acceptor Aldose Ribose 5-phosphate. *Biochemistry* 2007; **46**:12037-12052.

[52] Nikkola, M., Lindqvist, Y. and Schneider, G. Refined structure of transketolase from *Saccharomyces cerevisiae* at 2.0 A resolution. *J. Mol. Biol.* 1994; **238**:387-404.

[53] Gyamerah, M. and Willetts, A.J. Kinetics of overexpressed transketolase from *Escherichia coli* JM 107/pQR 700. *Enzyme Micro. Tech.* 1997; **20**:127-134.

[54] Nilsson, U., Meshalkina, L., Lindqvist, Y. and Schneider, G. Examination of Substrate Binding in Thiamin Diphosphate- dependent Transketolase by Protein Crystallography and Site-directed Mutagenesis. *J. Biol. Chem.* 1997; **272**:1864-1869.

[55] Nikkola, M., Lindqvist, Y. and Schneider, G. Refined Structure of Transketolase from *Saccharomyces cerevisiae* at 2·0 Å Resolution. *J. Mol. Biol.* 1994; **238**:387-404.

[56] Wikner, C., Nilsson, U., Meshalkina, L., Udekwu, C., Lindqvist, Y. and Schneider, G. Identification of Catalytically Important Residues in Yeast Transketolase. *Biochemistry* 1997; **36**:15643-15649.

[57] Fiedler, E., Thorell, S., Sandalova, T., Golbik, R., Konig, S. and Schneider, G. Snapshot of a key intermediate in enzymatic thiamin catalysis: crystal structure of the α-carbanion of (α,β-dihydroxyethyl)-thiamin diphosphate in the active site of transketolase from *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* 2002; **99**:591-595.

[58] Hibbert, E.G., Senussi, T., Costelloe, S.J., Lei, W., Smith, M.E., Ward, J.M., Hailes, H.C. and Dalby, P.A. Directed evolution of transketolase activity on non-phosphorylated substrates. *J. Biotechnol.* 2007; **131**:425-432.

[59] Hibbert, E.G., Senussi, T., Smith, M.E., Costelloe, S.J., Ward, J.M., Hailes, H.C. and Dalby, P.A. Directed evolution of transketolase substrate specificity towards an aliphatic aldehyde. *J. Biotechnol.* 2008; **134**:240-245.

[60] Ranoux, A., Karmee, S.K., Jin, J.F., Bhaduri, A., Caiazzo, A., Arends, I.W.C.E. and Hanefeld, U. Enhancement of the Substrate Scope of Transketolase. *ChemBioChem* 2012; **13**:1921-1931.

[61] Smith, M.E.B., Hibbert, E.G., Jones, A.B., Dalby, P.A. and Hailes, H.C. Enhancing and Reversing the Stereoselectivity of *Escherichia coli* Transketolase via Single-Point Mutations. *Adv. Synth. Catal.* 2008; **350**:2631-2638.

[62] Racker, E. Transketolase. In: *The Enzymes* (Eds P. Boyer, H. Lardy and K. Myrbäck). Academic Press, New York, 1961: pp. 397-406.

[63] Hailes, H.C., Dalby, P.A., Lye, G.J. and Ward, J.M. Biocatalytic approaches to ketodiols and aminodiols. *Chim. Oggi.* 2009; **27**:28-31.

[64] Cazares, A., Galman, J.L., Crago, L.G., Smith, M.E., Strafford, J., Rios-Solis, L., Lye, G.J., Dalby, P.A. and Hailes, H.C. Non-alpha-hydroxylated aldehydes with evolved transketolase enzymes. *Org. Biomol. Chem.* 2010; **8**:1301-1309.

[65] Abdoul-Zabar, J., Sorel, I., Hélaine, V., Charmantray, F., Devamani, T., Yi, D., Berardinis, d.V., Louis, D., Marlière, P., Fessner, W.-D. and Hecquet, L. Thermostable Transketolase from *Geobacillus stearothermophilus*: Characterization and Catalytic Properties. *Adv. Synth. Catal.* 2012; **In press**.

[66] Zimmermann, F.T., Schneider, A., Schorken, U., Sprenger, G.A. and Fessner, W.-D. Efficient multi-enzymatic synthesis of D-xylulose 5-phosphate. *Tetrahedron: Asymmetry* 1999; **10**:1643-1646.

[67] Horecker, B.L. and Smyrniotis, P.Z. Transketolase from liver and spinach: ribulose-5-P + ribose-5-P - sedoheptulose-7-P + glyceraldehyde-3-P. *Method. Enzymol.* 1955; **1**:371-375.

[68] Du, M.X., Sim, J., Fang, L., Zheng, Y., Koh, S., Stratton, J., Pons, J., Wang, J.J.-X. and Carte, B. Identification of novel small-molecule inhibitors for human transketolase by high-throughput screening with fluorescent intensity (FLINT) assay. *J. Biomol. Screening* 2004; **9**:427-433.

[69] Hecquet, L., Bolte, J. and Demuynck, C. New assays for transketolase. *Biosci. Biotech. Bioch.* 1993; **57**:2174-2176.

[70] Lee, J.-Y., Cheong, D.-E. and Kim, G.-J. A novel assay system for the measurement of transketolase activity using xylulokinase from *Saccharomyces cerevisiae*. *Biotechnol. Lett.* 2008; **30**:899-904.

[71] Naula, C., Alibu, V.P., Brock, J.M., Veitch, N.J., Burchmore, R.J.S. and Barrett, M.P. A new erythrose 4-phosphate dehydrogenase coupled assay for transketolase. *J. Biochem. Bioph. Meth.* 2008; **70**:1185-1187.

[72] Sevestre, A., Helaine, V., Guyot, G., Martin, C. and Hecquet, L. A fluorogenic assay for transketolase from Saccharomyces cerevisiae. *Tetrahedron Lett.* 2003; **44**:827-830.

[73] Sevestre, A., Charmantray, F., Helaine, V., Lasikova, A. and Hecquet, L. Synthesis of stereochemical probes for new fluorogenic assays for yeast transketolase variants. *Tetrahedron* 2006; **62**:3969-3976.

[74] Charmantray, F., Legeret, B., Helaine, V. and Hecquet, L. Fluorogenic substrates for the screening assay of transketolase through β-elimination of umbelliferone-Development, scope and limitations. *J. Biotechnol.* 2010; **145**:359-366.

[75] Kochetov, G.A., Usmanov, R.A. and Mevkh, A.T. A new method of determination of transketolase activity by asymmetric synthesis reaction. *Anal Biochem* 1978; **88**:296-301.

[76] Sevostyanova, I.A., Solovjeva, O.N. and Kochetov, G.A. Two methods for determination of transketolase activity. *Biochemistry (Moscow)* 2006; **71**:560-562.

[77] Kochetov, G.A. and Filippov, P.P. New method for continuous recording of the transketolase reaction. *Anal. Biochem.* 1972; **48**:286-291.

[78] Holldorf, A.W. Hydroxypyruvate. *Methods Enzym. Anal. (3rd Ed.)* 1984; **6**:578-582.

[79] Mitra, R.K. and Woodley, J.M. A useful assay for transketolase in asymmetric syntheses. *Biotechnol. Tech.* 1996; **10**:167-172.

[80] Miller, O.J., Hibbert, E.G., Ingram, C.U., Lye, G.J. and Dalby, P.A. Optimisation and evaluation of a generic microplate-based HPLC screen for transketolase activity. *Biotechnol. Lett.* 2007; **29**:1759-1770.

[81] Dische, Z. and Devi, A. A new colorimetric method for the determination of ketohexoses in presence of aldoses, ketoheptoses and ketopentoses. *Biochim. Biophys. Acta.* 1960; **39**:140-144.

[82] Smith, M.E.B., Kaulmann, U., Ward, J.M. and Hailes, H.C. A colorimetric assay for screening transketolase activity. *Bioorg. Med. Chem.* 2006; **14**:7062-7065.

[83] Janes, L.E., Lowendahl, A.C. and Kazlauskas, R.J. Quantitative screening of hydrolase libraries using pH indicators: identifying active and enantioselective hydrolases. *Chem. Eur. J.* 1998; **4**:2324-2331.

[84] Janes, L.E., Cimpoia, A. and Kazlauskas, R.J. Protease-Mediated Separation of Cis and Trans Diastereomers of 2(*R*,*S*)-benzyloxymethyl-4(*S*)-carboxylic Acid 1,3-Dioxolane Methyl Ester: Intermediates for the Synthesis of Dioxolane Nucleosides. *J. Org. Chem.* 1999; **64**:9019-9029.

[85] Somers, N.A. and Kazlauskas, R.J. Mapping the substrate selectivity and enantioselectivity of esterases from thermophiles. *Tetrahedron: Asymmetry* 2004; **15**:2991-3004.

[86] Chapman, E. and Wong, C.-H. A pH sensitive colorimetric assay for the high-throughput screening of enzyme inhibitors and substrates. A case study using kinases. *Bioorg. Med. Chem.* 2002; **10**:551-555.

[87] He, N., Yi, D. and Fessner, W.-D. Flexibility of Substrate Binding of Cytosine-5'-Monophosphate-N-Acetylneuraminate Synthetase (CMP-Sialate Synthetase) from *Neisseria meningitidis*: An Enabling Catalyst for the Synthesis of Neo-sialoconjugates. *Adv. Synth. Catal.* 2011; **353**:2384-2398.

[88] Deng, C. and Chen, R.R. A pH-sensitive assay for galactosyltransferase. *Anal. Biochem* 2004; **330**:219-226.

[89] Persson, M. and Palcic, M.M. A high-throughput pH indicator assay for screening glycosyltransferase saturation mutagenesis libraries. *Anal. Biochem.* 2008; **378**:1-7.

[90] Rosenberg, R.M., Herreid, R.M., Piazza, G.J. and O'Leary, M.H. Indicator assay for amino acid decarboxylases. *Anal. Biochem.* 1989; **181**:59-65.

[91] Hübner, G. and Golbik, R. A method for determination of transketolase activity based on the use of a pH indicator. *Biochem. Int.* 1992; **26**:545-550.

[92] Kazlauskas, R.J. Quantitative Assay of Hydrolases for Activity and Selectivity Using Color Changes. In: *Enzyme Assays. High-throughput Screening, Genetic Selection and Fingerprinting* (Ed J.-L. Reymond). Wiley-VCH, Weinheim, 2006: pp. 17-39.

[93] Hastings, A.B. and Sendroy, J., Jr. The effect of variation in ionic strength on the apparent first and second dissociation constants of carbonic acid. *J. Biol. Chem.* 1925; **65**:445-455.

[94] Sprenger, G.A., Schörken, U., Sprenger, G. and Sahm, H. Transketolase A of *Escherichia coli* K-12: Purification and properties of the enzyme from recombinant strains. *Eur. J. Biochem.* 1995; **230**:525-532.

[95] Hauptmann, S. and Gabler, W. Reaction of tris(hydroxymethyl)aminomethane with some aldehydes. *Zeitschrift fuer Naturforschung* 1968; **23**:111-112.

[96] Ogilvie, J.W. and Whitaker, S.C. Reaction of tris with aldehydes. Effect of tris on reactions catalyzed by homoserine dehydrogenase and glyceraldehyde-3-phosphate dehydrogenase. *Biochimica et Biophysica Acta, Enzymology* 1976; **445**:525-536.

[97] Bubb, W.A., Berthon, H.A. and Kuchel, P.W. Tris buffer reactivity with low-molecular-weight aldehydes: NMR characterization of the reactions of glyceraldehyde 3-phosphate. *Bioorg. Chem.* 1995; **23**:119-130.

[98] Smith, M.E.B., Smithies, K., Senussi, T., Dalby, P.A. and Hailes, H.C. The First Mimetic of the Transketolase Reaction. *European J. Org. Chem.*2006; **2006**:1121-1123.

[99] Rohr, K. and Mahrwald, R. Stereoselective Direct Amine-Catalyzed Decarboxylative Aldol Addition. *Org. Lett.* 2011; **13**:1878-1880.

[100] Long, G.L. and Winefordner, J.D. Limit of detection. A closer look at the IUPAC definition. *Anal. Chem.* 1983; **55**:712A-724A.

[101] AnalyticalMethodsCommittee Recommendations for the Definition, Estimation and Use of the Detection Limit. *Analyst* 1987; **112**:199-204.

[102] Kobori, Y., Myles, D.C. and Whitesides, G.M. Substrate specificity and carbohydrate synthesis using transketolase. *J. Org. Chem.* 1992; **57**:5899-5907.

[103] Hibbert, E.G., Senussi, T., Costelloe, S.J., Lei, W., Smith, M.E.B., Ward, J.M., Hailes, H.C. and Dalby, P.A. Directed evolution of transketolase activity on non-phosphorylated substrates. *J. Biotechnol.* 2007; **131**:425-432.

[104] Bolte, J., Demuynck, C. and Samaki, H. Utilization of enzymes in organic chemistry: transketolase catalyzed synthesis of ketoses. *Tetrahedron Lett.* 1987; **28**:5525-5528.

[105] Smith, M.E.B., Hibbert, E.G., Jones, A.B., Dalby, P.A. and Hailes, H.C. Enhancing and reversing the stereoselectivity of *Escherichia coli* transketolase via single-point mutations. *Adv. Synth. Catal.* 2008; **350**:2631-2638.

[106] Hibbert, E.G., Senussi, T., Smith, M.E.B., Costelloe, S.J., Ward, J.M., Hailes, H.C. and Dalby, P.A. Directed evolution of transketolase substrate specificity towards an aliphatic aldehyde. *J. Biotechnol.* 2008; **134**:240-245.

[107] Wahler, D., Badalassi, F., Crotti, P. and Reymond, J.-L. Enzyme fingerprints of activity, and stereo- and enantioselectivity from fluorogenic and chromogenic substrate arrays. *Chem. Eur. J.* 2002; **8**:3211-3228.

[108] Yi, D., Devamani, T., Abdoul-Zabar, J., Charmantray, F., Helaine, V., Hecquet, L. and Fessner, W.-D. A pH-Based High-Throughput Assay for Transketolase: Fingerprinting of Substrate Tolerance and Quantitative Kinetics. *ChemBioChem* 2012; **13**:2290-2300.

[109] Fiedler, E., Thorell, S., Sandalova, T., Golbik, R., König, S. and Schneider, G. Snapshot of a key intermediate in enzymatic thiamin catalysis: Crystal structure of the α-

carbanion of (α,β-dihydroxyethyl)-thiamin diphosphate in the active site of transketolase from *Saccharomyces cerevisiae*. *P. Natl. Acad. Sci. USA* 2002; **99**:591-595.

[110] Wikner, C., Nilsson, U., Meshalkina, L., Udekwu, C., Lindqvist, Y. and Schneider, G. Identification of catalytically important residues in yeast transketolase. *Biochemistry* 1997; **36**:15643-15649.

[111] Nilsson, U., Meshalkina, L., Lindqvist, Y. and Schneider, G. Examination of substrate binding in thiamin diphosphate-dependent transketolase by protein crystallography and site-directed mutagenesis. *J. Biol. Chem.* 1997; **272**:1864-1869.

[112] König, S., Schellenberger, A., Neef, H. and Schneider, G. Specificity of coenzyme binding in thiamin diphosphate-dependent enzymes. Crystal structures of yeast transketolase in complex with analogs of thiamin diphosphate. *J. Biol. Chem.* 1994; **269**:10879-10882.

[113] Kiefer, F., Arnold, K., Kunzli, M., Bordoli, L. and Schwede, T. The SWISS-MODEL Repository and associated resources. *Nucleic. Acids. Res.* 2009; **37**:D387-392.

[114] Arnold, K., Bordoli, L., Kopp, J. and Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 2006; **22**:195-201.

[115] Smith, M.E., Kaulmann, U., Ward, J.M. and Hailes, H.C. A colorimetric assay for screening transketolase activity. *Bioorg. Med. Chem.* 2006; **14**:7062-7065.

[116] Reetz, M.T. and Carballeira, J.D. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat. Protoc.* 2007; **2**:891-903.

[117] Hogrefe, H.H., Cline, J., Youngblood, G.L. and Allen, R.M. Creating randomized amino acid libraries with the QuikChange Multi Site-Directed Mutagenesis Kit. *Biotechniques* 2002; **33**:1158-1160, 1162, 1164-1155.

[118] An, Y., Chen, L., Sun, S., Lv, A. and Wu, W. QuikChange shuffling: a convenient and robust method for site-directed mutagenesis and random recombination of homologous genes. *N. Biotechno.l* 2011; **28**:320-325.

[119] Zheng, L., Baumann, U. and Reymond, J.-L. An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res.* 2004; **32**:e115.

[120] Zhang, J.-H., Chung, T.D.Y. and Oldenburg, K.R. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *J. Biomol. Screen* 1999; **4**:67-73.

[121] Niesen, F.H., Berglund, H. and Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc.* 2007; **2**:2212-2221.

[122] Ericsson, U.B., Hallberg, B.M., DeTitta, G.T., Dekker, N. and Nordlund, P. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal. Biochem.* 2006; **357**:289-298.

[123] Schauer, R. Chemistry, Metabolism, and Biological Functions of Sialic Acids. *Adv. Carbohyd. Chem. Bi.* 1982; **40**:131-234.

[124] Chen, X. and Varki, A. Advances in the Biology and Chemistry of Sialic Acids. *Acs Chem. Biol.* 2010; **5**:163-176.

[125] Varki, A. Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins. *Nature* 2007; **446**:1023-1029.

[126] Roy, R. and Baek, M.G. Glycodendrimers: novel glycotope isosteres unmasking sugar coding. case study with T-antigen markers from breast cancer MUC1 glycoprotein. *J. Biotechnol.* 2002; **90**:291-309.

[127] Lau, K., Yu, H., Thon, V., Khedri, Z., Leon, M.E., Tran, B.K. and Chen, X. Sequential two-step multienzyme synthesis of tumor-associated sialyl T-antigens and derivatives. *Org. Biomol. Chem.* 2011; **9**:2784-2789.

[128] Endo, T., Koizumi, S., Tabata, K., Kakita, S. and Ozaki, A. Large-scale production of the carbohydrate portion of the sialyl-Tn epitope, α-Neup5Ac-(2-->6)-D-GalpNAc, through bacterial coupling. *Carbohydr. Res.* 2001; **330**:439-443.

[129] Li, Y.H. and Chen, X. Sialic acid metabolism and sialyltransferases: natural functions and applications. *Appl. Microbiol. Biot.* 2012; **94**:887-905.

[130] Yu, H., Huang, S., Chokhawala, H., Sun, M., Zheng, H. and Chen, X. Highly efficient chemoenzymatic synthesis of naturally occurring and non-natural α-2,6-linked sialosides: a *P. damsela* α-2,6-sialyltransferase with extremely flexible donor-substrate specificity. *Angew. Chem. Int. Ed. Engl.* 2006; **45**:3938-3944.

[131] Yu, H., Cheng, J.S., Ding, L., Khedri, Z., Chen, Y., Chin, S., Lau, K., Tiwari, V.K. and Chen, X. Chemoenzymatic Synthesis of GD3 Oligosaccharides and Other Disialyl Glycans Containing Natural and Non-natural Sialic Acids. *J. Am. Chem. Soc.* 2009; **131**:18467-18477.

[132] Cao, H.Z., Muthana, S., Li, Y.H., Cheng, J.S. and Chen, X. Parallel chemoenzymatic synthesis of sialosides containing a C5-diversified sialic acid. *Bioorg. Med. Chem. Lett.* 2009; **19**:5869-5871.

[133] Khedri, Z., Muthana, M.M., Li, Y.H., Muthana, S.M., Yu, H., Cao, H.Z. and Chen, X. Probe sialidase substrate specificity using chemoenzymatically synthesized sialosides containing C9-modified sialic acid. *Chem. Commun.* 2012; **48**:3357-3359.

[134] Cao, H.Z., Li, Y.H., Lau, K., Muthana, S., Yu, H., Cheng, J.S., Chokhawala, H.A., Sugiarto, G., Zhang, L. and Chen, X. Sialidase substrate specificity studies using chemoenzymatically synthesized sialosides containing C5-modified sialic acids. *Organic & Biomolecular Chemistry* 2009; **7**:5137-5145.

[135] Morley, T.J. and Withers, S.G. Chemoenzymatic synthesis and enzymatic analysis of 8-modified cytidine monophosphate-sialic acid and sialyl lactose derivatives. *J. Am. Chem. Soc.* 2010; **132**:9430-9437.

[136] Yu, H., Cao, H.Z., Tiwari, V.K., Li, Y.H. and Chen, X. Chemoenzymatic synthesis of C8-modified sialic acids and related alpha 2-3-and alpha 2-6-linked sialosides. *Bioorganic & Medicinal Chemistry Lett.* 2011; **21**:5037-5040.

[137] Sugiarto, G., Lau, K., Qu, J., Li, Y., Lim, S., Mu, S., Ames, J.B., Fisher, A.J. and Chen, X. A Sialyltransferase Mutant with Decreased Donor Hydrolysis and Reduced Sialidase Activities for Directly Sialylating Lewisx. *ACS Chem. Biol.* 2012; **7**:1232-1240.

[138] Vandijk, W. and Vandeneijnden, D.H. Cytidine Monophosphate N-Acetylneuraminate Synthetase - Nuclear Enzyme. *Biochem. Soc. T.* 1977; **5**:60-60.

[139] Mizanur, R.M. and Pohl, N.L. Bacterial CMP-sialic acid synthetases: production, properties, and applications. *Appl. Microbiol. Biot.* 2008; **80**:757-765.

[140] Li, Y.H., Yu, H., Cao, H.Z., Muthana, S. and Chen, X. Pasteurella multocida CMP-sialic acid synthetase and mutants of *Neisseria meningitidis* CMP-sialic acid synthetase with improved substrate promiscuity. *Appl. Microbiol. Biot.* 2012; **93**:2411-2423.

[141] Knorst, M. and Fessner, W.D. CMP-sialate synthetase from *Neisseria meningitidis* - Overexpression and application to the synthesis of oligosaccharides containing modified sialic acids. *Adv. Synth. Catal.* 2001; **343**:698-710.

[142] Mosimann, S.C., Gilbert, M., Dombrowski, D., To, R., Wakarchuk, W. and Strynadka, N.C.J. Structure of a sialic acid-activating synthetase, CMP-acylneuraminate synthetase in the presence and absence of CDP. *J. Biol. Chem.* 2001; **276**:8190-8196.

[143] He, N., Yi, D. and Fessner, W.D. Flexibility of Substrate Binding of Cytosine-5 '-Monophosphate-N-Acetylneuraminate Synthetase (CMP-Sialate Synthetase) from *Neisseria meningitidis*: An Enabling Catalyst for the Synthesis of Neo-sialoconjugates. *Adv. Synth. Catal.* 2011; **353**:2384-2398.

[144] Cheng, J.S., Huang, S.S., Yu, H., Li, Y.H., Lau, K. and Chen, X. Trans-sialidase activity of *Photobacterium damsela* α 2,6-sialyltransferase and its application in the synthesis of sialosides. *Glycobiology* 2010; **20**:260-268.

[145] Yamamoto, T., Nakashizuka, M. and Terada, I. Cloning and expression of a marine bacterial beta-galactoside α2,6-sialyltransferase gene from *Photobacterium damsela* JT0160. *J. Biochem.* 1998; **123**:94-100.

[146] Kajihara, Y., Kamitani, T. and Sakakibara, T. A new fluorescent assay for sialyltransferase. *Carbohydr. Re.s* 2001; **331**:455-459.

[147] Yu, H., Chokhawala, H., Karpel, R., Wu, B., Zhang, J., Zhang, Y., Jia, Q. and Chen, X. A multifunctional Pasteurella multocida sialyltransferase: a powerful tool for the synthesis of sialoside libraries. *J. Am. Chem. Soc.* 2005; **127**:17618-17619.

[148] Iwatani, T., Okino, N., Sakakura, M., Kajiwara, H., Takakura, Y., Kimura, M., Ito, M., Yamamoto, T. and Kakuta, Y. Crystal structure of α/β-galactoside α2,3-sialyltransferase from a luminous marine bacterium, *Photobacterium phosphoreum*. *FEBS Lett.* 2009; **583**:2083-2087.

[149] Kakuta, Y., Okino, N., Kajiwara, H., Ichikawa, M., Takakura, Y., Ito, M. and Yamamoto, T. Crystal structure of Vibrionaceae *Photobacterium* sp. JT-ISH-224 α2,6-sialyltransferase in a ternary complex with donor product CMP and acceptor substrate lactose: catalytic mechanism and substrate recognition. *Glycobiology* 2008; **18**:66-73.

[150] Ni, L., Chokhawala, H.A., Cao, H., Henning, R., Ng, L., Huang, S., Yu, H., Chen, X. and Fisher, A.J. Crystal structures of Pasteurella multocida sialyltransferase complexes with acceptor and donor analogues reveal substrate binding sites and catalytic mechanism. *Biochemistry* 2007; **46**:6288-6298.

[151] Gilbert, M., Brisson, J.R., Karwaski, M.F., Michniewicz, J., Cunningham, A.M., Wu, Y.Y., Young, N.M. and Wakarchuk, W.W. Biosynthesis of ganglioside mimics in

*Campylobacter jejuni* OH4384 - Identification of the glycosyltransferase genes, enzymatic synthesis of model compounds, and characterization of nanomole amounts by 600-MHz H-1 and C-13 NMR analysis. *J. Biol. Chem.* 2000; **275**:3896-3906.

[152] Fox, K.L., Cox, A.D., Gilbert, M., Wakarchuk, W.W., Li, J., Makepeace, K., Richards, J.C., Moxon, E.R. and Hood, D.W. Identification of a bifunctional lipopolysaccharide sialyltransferase in *Haemophilus influenzae*: incorporation of disialic acid. *J. Biol. Chem.* 2006; **281**:40024-40032.

[153] Jones, P.A., Samuels, N.M., Phillips, N.J., Munson, R.S., Jr., Bozue, J.A., Arseneau, J.A., Nichols, W.A., Zaleski, A., Gibson, B.W. and Apicella, M.A. *Haemophilus influenzae* type b strain A2 has multiple sialyltransferases involved in lipooligosaccharide sialylation. *J. Biol. Chem.* 2002; **277**:14598-14611.

[154] Li, Y., Sun, M., Huang, S., Yu, H., Chokhawala, H.A., Thon, V. and Chen, X. The Hd0053 gene of *Haemophilus ducreyi* encodes an α2,3-sialyltransferase. *Biochem. Biophys. Res. Commun.* 2007; **361**:555-560.

[155] Gilbert, M., Watson, D.C., Cunningham, A.M., Jennings, M.P., Young, N.M. and Wakarchuk, W.W. Cloning of the lipooligosaccharide α-2,3-sialyltransferase from the bacterial pathogens *Neisseria meningitidis* and *Neisseria gonorrhoeae*. *J. Biol. Chem.* 1996; **271**:28271-28276.

[156] Thon, V., Lau, K., Yu, H., Tran, B.K. and Chen, X. PmST2: A novel *Pasteurella multocida* glycolipid α 2-3-sialyltransferase. *Glycobiology* 2011; **21**:1206-1216.

[157] Thon, V., Li, Y.H., Yu, H., Lau, K. and Chen, X. PmST3 from *Pasteurella multocida* encoded by Pm1174 gene is a monofunctional α-2,3-sialyltransferase. *Appl. Microbiol. Biot.* 2012; **94**:977-985.

[158] Yamamoto, T., Nakashizuka, M., Kodama, H., Kajihara, Y. and Terada, I. Purification and characterization of a marine bacterial β-galactoside alpha 2,6-sialyltransferase from *Photobacterium damsela* JT0160. *J. Biochem.* 1996; **120**:104-110.

[159] Sun, M., Li, Y., Chokhawala, H.A., Henning, R. and Chen, X. N-Terminal 112 amino acid residues are not required for the sialyltransferase activity of *Photobacterium damsela* α2,6-sialyltransferase. *Biotechnol Lett* .2008; **30**:671-676.

[160] Yamamoto, T., Hamada, Y., Ichikawa, M., Kajiwara, H., Mine, T., Tsukamoto, H. and Takakura, Y. A β-galactoside α2,6-sialyltransferase produced by a marine bacterium, *Photobacterium leiognathi* JT-SHIZ-145, is active at pH 8. *Glycobiology* 2007; **17**:1167-1174.

[161] Tsukamoto, H., Takakura, Y. and Yamamoto, T. Purification, cloning, and expression of an α/β-galactoside α-2,3-sialyltransferase from a luminous marine bacterium, *Photobacterium phosphoreum*. *J. Biol. Chem.* 2007; **282**:29794-29802.

[162] Tsukamoto, H., Takakura, Y., Mine, T. and Yamamoto, T. *Photobacterium* sp. JT-ISH-224 produces two sialyltransferases, α-/β-galactoside α2,3-sialyltransferase and β-galactoside α2,6-sialyltransferase. *J. Biochem.* 2008; **143**:187-197.

[163] Takakura, Y., Tsukamoto, H. and Yamamoto, T. Molecular cloning, expression and properties of an α/β-Galactoside α2,3-sialyltransferase from *Vibrio* sp. JT-FAJ-16. *J. Biochem.* 2007; **142**:403-412.

[164] Park, S., Lee, M.R. and Shin, I. Carbohydrate microarrays as powerful tools in studies of carbohydrate-mediated biological processes. *Chem. Commun.* 2008; 4389-4399.

[165] Blixt, O., Allin, K., Bohorov, O., Liu, X.F., Andersson-Sand, H., Hoffmann, J. and Razi, N. Glycan microarrays for screening sialyltransferase specificities. *Glycoconjugate J.* 2008; **25**:59-68.

[166] Blixt, O., Han, S.F., Liao, L., Zeng, Y., Hoffmann, J., Futakawa, S. and Paulson, J.C. Sialoside analogue arrays for rapid identification of high affinity siglec ligands. *J. Am. Chem. Soc.* 2008; **130**:6680.

[167] Song, X.Z., Yu, H., Chen, X., Lasanajak, Y., Tappert, M.M., Air, G.M., Tiwari, V.K., Cao, H.Z., Chokhawala, H.A., Zheng, H.J., Cummings, R.D. and Smith, D.F. A Sialylated Glycan Microarray Reveals Novel Interactions of Modified Sialic Acids with Proteins and Viruses. *J. Biol. Chem.* 2011; **286**:31610-31622.

[168] Linman, M.J., Yu, H., Chen, X. and Cheng, Q. Fabrication and Characterization of a Sialoside-Based Carbohydrate Microarray Biointerface for Protein Binding Analysis with Surface Plasmon Resonance Imaging. *Acs Appl. Mater. Inter.* 2009; **1**:1755-1762.

[169] Heyl, D., Rikowski, E., Hoffmann, R.C., Schneider, J.J. and Fessner, W.D. A "Clickable" Hybrid Nanocluster of Cubic Symmetry. *Chem-Eur. J.* 2010; **16**:5544-5548.

[170] Boas, U. and Heegaard, P.M.H. Dendrimers in drug research. *Chem. Soc. Rev.* 2004; **33**:43-63.

[171] He, N., Yi, D. and Fessner, W.-D. Flexibility of Substrate Binding of Cytosine-5′-Monophosphate-*N*-Acetylneuraminate Synthetase (CMP-Sialate Synthetase) from *Neisseria meningitidis*: An Enabling Catalyst for the Synthesis of Neo-sialoconjugates. *Adv. Synth. Catal.* 2011; **353**:2384-2398.

[172] Yu, H. and Chen, X. Carbohydrate post-glycosylational modifications. *Org. Biomol. Chem.* 2007; **5**:865-872.

[173] Lau, K., Thon, V., Yu, H., Ding, L., Chen, Y., Muthana, M.M., Wong, D., Huang, R. and Chen, X. Highly efficient chemoenzymatic synthesis of β1-4-linked galactosides with promiscuous bacterial β1-4-galactosyltransferases. *Chem. Commun. (Camb)* 2010; **46**:6066-6068.

[174] Wakarchuk, W.W., Cunningham, A., Watson, D.C. and Young, N.M. Role of paired basic residues in the expression of active recombinant galactosyltransferases from the bacterial pathogen *Neisseria meningitidis*. *Protein Eng.* 1998; **11**:295-302.

[175] Wakarchuk, W.W., Cunningham, A., Watson, D.C. and Young, N.M. Role of paired basic residues in the expression of active recombinant galactosyltransferases from the bacterial pathogen *Neisseria meningitidis*. *Protein Eng.*1998; **11**:295-302.

[176] Park, J.E., Lee, K.Y., Do, S.I. and Lee, S.S. Expression and characterization of β-1,4-galactosyltransferase from *Neisseria meningitidis* and *Neisseria gonorrhoeae*. *J. Biochem. Mol. Biol.* 2002; **35**:330-336.

[177] Endo, T., Koizumi, S., Tabata, K. and Ozaki, A. Cloning and expression of β 1,4-galactosyltransferase gene from *Helicobacter pylori*. *Glycobiology* 2000; **10**:809-813.

[178] Logan, S.M., Conlan, J.W., Monteiro, M.A., Wakarchuk, W.W. and Altman, E. Functional genomics of *Helicobacter pylori*: identification of a β-1,4 galactosyltransferase and generation of mutants with altered lipopolysaccharide. *Mol. Microbiol.* 2000; **35**:1156-1167.

[179] Chien, W.T., Liang, C.F., Yu, C.C., Lin, J.H., Wu, H.T. and Lin, C.C. Glucose 1-Phosphate Thymidylyltransferase in the Synthesis of Uridine 5 '-Diphosphate Galactose and its Application in the Synthesis of *N*-Acetyllactosamine. *Adv. Synth. Catal.* 2012; **354**:123-132.

[180] Dumon, C., Bosso, C., Utille, J.P., Heyraud, A. and Samain, E. Production of Lewis x tetrasaccharides by metabolically engineered *Escherichia coli*. *ChemBioChem* 2006; **7**:359-365.

[181] Priem, B., Gilbert, M., Wakarchuk, W.W., Heyraud, A. and Samain, E. A new fermentation process allows large-scale production of human milk oligosaccharides by metabolically engineered bacteria. *Glycobiology* 2002; **12**:235-240.

[182] Endo, T., Koizumi, S., Tabata, K., Kakita, S. and Ozaki, A. Large-scale production of *N*-acetyllactosamine through bacterial coupling. *Carbohydrate Res.* 1999; **316**:179-183.

[183] Koizumi, S., Endo, T., Tabata, K. and Ozaki, A. Large-scale production of UDP-galactose and globotriose by coupling metabolically engineered bacteria. *Nat. Biotechnol.* 1998; **16**:847-850.

[184] Lau, K., Thon, V., Yu, H., Ding, L., Chen, Y., Muthana, M.M., Wong, D., Huang, R. and Chen, X. Highly efficient chemoenzymatic synthesis of 1-4-linked galactosides with promiscuous bacterial β1-4-galactosyltransferases. *Chem. Commun.* 2010; **46**:6066-6068.

[185] Namdjou, D.J., Chen, H.M., Vinogradov, E., Brochu, D., Withers, S.G. and Wakarchuk, W.W. A β1,4-galactosyltransferase from *Helicobacter pylori* is an efficient and versatile biocatalyst displaying a novel activity for thioglycoside synthesis. *ChemBioChem* 2008; **9**:1632-1640.

[186] Sauerzapfe, B., Krenek, K., Schmiedel, J., Wakarchuk, W.W., Pelantova, H., Kren, V. and Elling, L. Chemo-enzymatic synthesis of poly-*N*-acetyllactosamine (poly-LacNAc) structures and their characterization for CGL2-galectin-mediated binding of ECM glycoproteins to biomaterial surfaces. *Glycoconjugate Journal* 2009; **26**:141-159.

[187] Chen, X., Kowal, P., Hamad, S., Fan, H. and Wang, P.G. Cloning, expression and characterization of a UDP-galactose 4-epimerase from *Escherichia coli*. *Biotechnol. Lett.* 1999; **21**:1131-1135.

[188] Blixt, O., Brown, J., Schur, M.J., Wakarchuk, W. and Paulson, J.C. Efficient preparation of natural and synthetic galactosides with a recombinant β-1,4-galactosyltransferase-/UDP-4'-gal epimerase fusion protein. *J. Org. Chem.* 2001; **66**:2442-2448.

[189] Bettler, E., Samain, E., Chazalet, V., Bosso, C., Heyraud, A., Joziasse, D.H., Wakarchuk, W.W., Imberty, A. and Geremia, R.A. The living factory: *In vivo* production of *N*-acetyllactosamine containing carbohydrates in *E-coli*. *Glycoconjugate J.*1999; **16**:205-212.

[190] Logan, S.M., Conlan, J.W., Monteiro, M.A., Wakarchuk, W.W. and Altman, E. Functional genomics of Helicobacter pylori: identification of a β-1,4 galactosyltransferase and generation of mutants with altered lipopolysaccharide. *Mol. Microb.* 2000; **35**:1156-1167.

[191] Thomas, J.G., Ayling, A. and Baneyx, F. Molecular chaperones, folding catalysts, and the recovery of active recombinant proteins from *E. coli*. To fold or to refold. *Appl. Biochem. Biotechnol* .1997; **66**:197-238.

[192] Nishihara, K., Kanemori, M., Yanagi, H. and Yura, T. Overexpression of trigger factor prevents aggregation of recombinant proteins in *Escherichia coli*. *Appl. Environ. Microbiol.* 2000; **66**:884-889.

[193] Nishihara, K., Kanemori, M., Kitagawa, M., Yanagi, H. and Yura, T. Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in *Escherichia coli*. *Appl. Environ. Microbiol.* 1998; **64**:1694-1699.

[194] Naruchi, K. and Nishimura, S.-I. Membrane-Bound Stable Glycosyltransferases: Highly Oriented Protein Immobilization by a C-Terminal Cationic Amphipathic Peptide. *Angew. Chem. Int. Edit.* 2011; **50**:1328-1331.

[195] Ma, B., Wang, G., Palcic, M.M., Hazes, B. and Taylor, D.E. C-terminal amino acids of *Helicobacter pylori* α1,3/4 fucosyltransferases determine type I and type II transfer. *J. Biol. Chem.* 2003; **278**:21893-21900.

[196] Unruh, P. Dissertation, Technische Universität Darmstadt. 2007.

[197] Michael, G.B., Kadlec, K., Sweeney, M.T., Brzuszkiewicz, E., Liesegang, H., Daniel, R., Murray, R.W., Watts, J.L. and Schwarz, S. ICEPmu1, an integrative conjugative element (ICE) of *Pasteurella multocida*: structure and transfer. *J. Antimicrob. Chemoth.* 2012; **67**:91-100.

[198] Liu, W., Yang, M., Xu, Z., Zheng, H., Liang, W., Zhou, R., Wu, B. and Chen, H. Complete Genome Sequence of *Pasteurella multocida* HN06, a Toxigenic Strain of Serogroup D. *J. Bacteriol.* 2012; **194**:3292-3293.

[199] May, B.J., Zhang, Q., Li, L.L., Paustian, M.L., Whittam, T.S. and Kapur, V. Complete genomic sequence of *Pasteurella multocida*, Pm70. *P. Natl. A. of Sci.* 2001; **98**:3460-3465.

[200] Mine, T., Katayama, S., Kajiwara, H., Tsunashima, M., Tsukamoto, H., Takakura, Y. and Yamamoto, T. An α2,6-sialyltransferase cloned from *Photobacterium leiognathi* strain JT-SHIZ-119 shows both sialyltransferase and neuraminidase activity. *Glycobiology* 2010; **20**:158-165.

[201] Wang, F., Wang, J., Jian, H., Zhang, B., Li, S., Zeng, X., Gao, L., Bartlett, D.H., Yu, J., Hu, S. and Xiao, X. Environmental adaptation: genomic analysis of the piezotolerant and psychrotolerant deep-sea iron reducing bacterium *Shewanella piezotolerans* WP3. *PLoS ONE* 2008; **3**:e1937.

[202] Yu, H., Chokhawala, H.A., Huang, S. and Chen, X. One-pot three-enzyme chemoenzymatic approach to the synthesis of sialosides containing natural and non-natural functionalities. *Nat. Protoc.* 2006; **1**:2485-2492.

[203] Ni, L., Sun, M., Yu, H., Chokhawala, H., Chen, X. and Fisher, A.J. Cytidine 5'-monophosphate (CMP)-induced structural changes in a multifunctional sialyltransferase from *Pasteurella multocida*. *Biochemistry* 2006; **45**:2139-2148.

[204] Huang, S.S., Yu, H. and Chen, X. Chemoenzymatic synthesis of α2-3-sialylated carbohydrate epitopes. *Sci. China. Chem.* 2011; **54**:117-128.

[205] Muthana, S., Yu, H., Cao, H.Z., Cheng, J.S. and Chen, X. Chemoenzymatic Synthesis of a New Class of Macrocyclic Oligosaccharides. *J. Org. Chem.* 2009; **74**:2928-2936.

[206] Sabesan, S. and Paulson, J.C. Combined chemical and enzymatic synthesis of sialyloligosaccharides and characterization by 500-MHz and proton and carbon-13 NMR spectroscopy. *J. Am. Chem. Soc.* 1986; **108**:2068-2080.

[207] Deng, C. and Chen, R.R. A pH-sensitive assay for galactosyltransferase. *Anal. Biochem* .2004; **330**:219-226.

[208] Persson, M. and Palcic, M.M. A high-throughput pH indicator assay for screening glycosyltransferase saturation mutagenesis libraries. *Anal. Biochem.* 2008; **378**:1-7.

[209] Chapman, E. and Wong, C.H. A pH sensitive colorometric assay for the high-throughput screening of enzyme inhibitors and substrates: a case study using kinases. *Bioorg. Med. Chem.* 2002; **10**:551-555.

[210] Okino, N., Kakuta, Y., Kajiwara, H., Ichikawa, M., Takakura, Y., Ito, M. and Yamamoto, T. Purification, crystallization and preliminary crystallographic characterization of the α 2,6-sialyltransferase from *Photobacterium* sp. JT-ISH-224. *Acta. Crystallogr. Sect. F. Struct. Biol. Cryst. Commun.* 2007; **63**:662-664.

[211] Schauer, R. Achievements and challenges of sialic acid research. *Glycoconjugate J.*2000; **17**:485-499.

[212] Carlin, A.F., Uchiyama, S., Chang, Y.-C., Lewis, A.L., Nizet, V. and Varki, A. Molecular mimicry of host sialylated glycans allows a bacterial pathogen to engage neutrophil Siglec-9 and dampen the innate immune response. *Blood* 2009; **113**:3333-3336.

[213] Mandrell, R.E. and Apicella, M.A. Lipo-oligosaccharides (LOS) of mucosal pathogens: molecular mimicry and host-modification of LOS. *Immunobiology* 1993; **187**:382-402.

[214] Oschlies, M., Dickmanns, A., Haselhorst, T., Schaper, W., Stummeyer, K., Tiralongo, J., Weinhold, B., Gerardy-Schahn, R., von Itzstein, M., Ficner, R. and Münster-Kühnel, A.-K. A C-Terminal Phosphatase Module Conserved in Vertebrate CMP-Sialic Acid Synthetases Provides a Tetramerization Interface for the Physiologically Active Enzyme. *J. Mol. Biol.* 2009; **393**:83-97.

[215] Krapp, S., Münster-Kühnel, A.K., Kaiser, J.T., Huber, R., Tiralongo, J., Gerardy-Schahn, R. and Jacob, U. The Crystal Structure of Murine CMP-5-N-acetylneuraminic Acid Synthetase. *J. Mol. Biol.* 2003; **334**:625-637.

[216] Horsfall, L.E., Nelson, A. and Berry, A. Identification and characterization of important residues in the catalytic mechanism of CMP-Neu5Ac synthetase from *Neisseria meningitidis*. *FEBS J.* 2010; **277**:2779-2790.

[217] Reetz, M.T. and Carballeira, J.D. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat. protoc.* 2007; **2**:891-903.

[218] Coutinho, P.M., Deleury, E., Davies, G.J. and Henrissat, B. An evolving hierarchical family classification for glycosyltransferases. *J. Mol. Biol.* 2003; **328**:307-317.

[219] Fukae, K., Yamamoto, N., Hatakeyama, Y. and Kajihara, Y. Chemoenzymatic synthesis of diverse asparagine-linked α-(2,3)-sialyloligosaccharides. *Glycoconj. J.* 2004; **21**:243-250.

[220] Kajihara, Y., Akai, S., Nakagawa, T., Sato, R., Ebata, T., Kodama, H. and Sato, K. Enzymatic synthesis of Kdn oligosaccharides by a bacterial α-(2-6)-sialyltransferase. *Carbohydr. Res.* 1999; **315**:137-141.

[221] Ding, L., Yu, H., Lau, K., Li, Y.H., Muthana, S., Wang, J.R. and Chen, X. Efficient chemoenzymatic synthesis of sialyl Tn-antigens and derivatives. *Chem. Commun.* 2011; **47**:8691-8693.

[222] Yamamoto, T., Nagae, H., Kajihara, Y. and Terada, I. Mass production of bacterial α 2,6-sialyltransferase and enzymatic syntheses of sialyloligosaccharides. *Biosci. Biotechnol. Biochem.* 1998; **62**:210-214.

[223] Rohfritsch, P.F., Joosten, J.A., Krzewinski-Recchi, M.A., Harduin-Lepers, A., Laporte, B., Juliant, S., Cerutti, M., Delannoy, P., Vliegenthart, J.F. and Kamerling, J.P. Probing the substrate specificity of four different sialyltransferases using synthetic β-D-Galp-(1-->4)-β-D-GlcpNAc-(1-->2)-α-D-Manp-(1-->O) $(CH_2)_7CH_3$ analogues general activating effect of replacing *N*-acetylglucosamine by *N*-propionylglucosamine. *Biochim. Biophys. Acta.* 2006; **1760**:685-692.

[224] Mine, T., Miyazaki, T., Kajiwara, H., Tateda, N., Ajisaka, K. and Yamamoto, T. A recombinant α-(2-3)-sialyltransferase with an extremely broad acceptor substrate specificity from *Photobacterium* sp. JT-ISH-224 can transfer *N*-acetylneuraminic acid to inositols. *Carbohydr. Res.* 2010; **345**:2485-2490.

[225] Mine, T., Miyazaki, T., Kajiwara, H., Naito, K., Ajisaka, K. and Yamamoto, T. Enzymatic synthesis of unique sialyloligosaccharides using marine bacterial α-(2-3)- and α-(2-6)-sialyltransferases. *Carbohydr. Res.* 2010; **345**:1417-1421.

[226] Wang, F., Wang, P., Chen, M. and Xiao, X. Isolation of extremophiles with the detection and retrieval of Shewanella strains in deep-sea sediments from the west Pacific. *Extremophiles* 2004; **8**:165-168.

[227] Feller, G., Arpigny, J.L., Narinx, E. and Gerday, C. Molecular adaptations of enzymes from psychrophilic organisms. *Compar. Biochem. Physiol. P-A Physiol.* 1997; **118**:495-499.

[228] Feller, G., Payan, F., Theys, F., Qian, M., Haser, R. and Gerday, C. Stability and structural analysis of α-amylase from the antarctic psychrophile *Alteromonas haloplanctis* A23. *Eur. J. Biochem.* 1994; **222**:441-447.

[229] Margesin, R., Palma, N., Knauseder, F. and Schinner, F. Proteases of psychrotrophic bacteria isolated from glaciers. *Journal of Basic Microbiology* 1991; **31**:377-383.

[230] Wang, F., Xiao, X., Ou, H.Y. and Gai, Y. Role and regulation of fatty acid biosynthesis in the response of *Shewanella piezotolerans* WP3 to different temperatures and pressures. *J. Bacteriol.* 2009; **191**:2574-2584.

[231] Li, S., Xiao, X., Sun, P. and Wang, F. Screening of genes regulated by cold shock in Shewanella piezotolerans WP3 and time course expression of cold-regulated genes. *Arch. Microbiol.* 2008; **189**:549-556.

[232] Xiao, X., Wang, P., Zeng, X., Bartlett, D.H. and Wang, F. Shewanella psychrophila sp. nov. and *Shewanella piezotolerans* sp. nov., isolated from west Pacific deep-sea sediment. *Int. J. Syst. Evol. Microbiol.* 2007; **57**:60-65.

[233] Wang, F., Li, Q. and Xiao, X. A novel filamentous phage from the deep-sea bacterium *Shewanella piezotolerans* WP3 is induced at low temperature. *J. Bacteriol.* 2007; **189**:7151-7153.

[234] Jones, P.G., VanBogelen, R.A. and Neidhardt, F.C. Induction of proteins in response to low temperature in *Escherichia coli*. *J. Bacteriol.* 1987; **169**:2092-2095.

[235] Ferrer, M., Chernikova, T.N., Yakimov, M.M., Golyshin, P.N. and Timmis, K.N. Chaperonins govern growth of *Escherichia coli* at low temperatures. *Nat. Biotechnol.* 2003; **21**:1266-1267.

[236] Juretzek, T., Le Dall, M., Mauersberger, S., Gaillardin, C., Barth, G. and Nicaud, J. Vectors for gene expression and amplification in the yeast *Yarrowia lipolytica*. *Yeast* 2001; **18**:97-113.

[237] Barth, G. and Gaillardin, C. Physiology and genetics of the dimorphic fungus *Yarrowia lipolytica*. *FEMS Microbiol. Rev.* 1997; **19**:219-237.

[238] Otto, C., Yovkova, V., Aurich, A., Mauersberger, S. and Barth, G. Variation of the by-product spectrum during alpha-ketoglutaric acid production from raw glycerol by overexpression of fumarase and pyruvate carboxylase genes in *Yarrowia lipolytica*. *Appl. Microbiol. Biotechnol.* 2012; **95**:905-917.

## List of Schemes, Figures, Tables and Sequences

### Schemes



### Figures

## Tables

## Sequences

**Curriculum Vitae**


# Yi, Dong

**ADDRESS:** Wilhelm-Jäger-Str. 10
64287, Darmstadt
Germany

**EMAIL:** charlesdyi@yahoo.com


**DATE:** Oct, 2012

## PERSONAL INFORMATION

Birthday: 15.11.1981

Birthplace: Shaoxing, Zhejiang Province, P. R. China

Nationality: Chinese

## EDUCATION:

| | |
|---|---|
| 04.2009- | Technische Universität Darmstadt, Darmstadt, Germany<br>Ph.D. candidate |
| 09.2005-06.2008 | Shanghai Institute of Pharmaceutical Industry, Shanghai, P. R. China<br>Master of Medicine in Microbial & Biochemical Pharmaceutics |
| 09.2000-06.2004 | China Pharmaceutical University, Nanjing, P. R. China<br>Bachelor of Engineering in Bioengineering |
| 09.1997-06.2000 | Shaoxing First High School, Shaoxing, P. R. China |
| 09.1994-06.1997 | Fushan Middle School, Shaoxing, P. R. China |
| 09.1988-06.1994 | Shengli Primary School, Shaoxing, P. R. China |

## WORK EXPERIENCE:

| | |
|---|---|
| 06.2008-09.2008 | Shanghai Institute of Pharmaceutical Industry,<br>Research associate |

## Publications

**Dong Yi**, Titu Devamani, Wolf-Dieter Fessner *et al*., A pH-Based High-Throughput Assay for Transketolase — Fingerprinting of Substrate Tolerance and Quantitative Kinetics. *ChemBioChem*, 2012, 13(15), 2290-2300.

Ning He, **Dong Yi** and Wolf-Dieter Fessner, Flexibility of Substrate Binding of Cytosine-5'-Monophosphate-*N*-Acetylneuraminate Synthetase (CMP-Sialate Synthetase) from *Neisseria meningitidis*: An Enabling Catalyst for the Synthesis of Neo-sialoconjugates. *Advanced Synthesis & Catalysis*, 2011, 353(13), 2384-2398.

Juliane Abdoul Zabar, Isabelle Sorel, Virgil Hélaine, Franck Charmantray, Titu Devamani, **Dong Yi** *et al.*, Thermostable Transketolase from *Geobacillus stearothermophilus*: Characterization and Catalytic Properties. *Advanced Synthesis & Catalysis*, 2012. (in press)

## Conference Posters

**Dong Yi**, Titu Devamania, Isabelle Sorel, Laurence Hecquet and Wolf-Dieter Fessner, Transketolase Catalysis for Asymmetric Carboligation: Directed Evolution of Substrate Scope and Stereoselectivity. *Protein Engineering for Biocatalysis*, Greifswald, Germany, 2012.

**Dong Yi**, Titu Devamani and Wolf-Dieter Fessner, Transketolase catalyzing asymmetric carboligation: Directed evolution of substrate scope towards potential industrial applications. *Catalyzing Bio-Economy – Biocatalysts for Industrial Biotechnology*, DECHEMA, Frankfurt am Main, Germany, 2012.

**Dong Yi**, Ning He and Wolf-Dieter Fessner, Chemoenzymatic Synthesis of neo-Sialoconjugates Using Bacterial Sialyltransferases. *Biotrans International Conference*, Giardini Naxos/Sicily, Italy, 2011.

**Dong Yi**, Titu Devamani and Wolf-Dieter Fessner, A pH-Based High-Throughput Assay for Transketolase. *COST-Training School*, University of Siena, Italy, 2011.

Yi, Dong
Wilhelm-Jäger-Str. 10
64287, Darmstadt 31.10.2012

# Erklärung

Ich erkläre hiermit, daß ich meine Dissertation selbständig und nur mit den angegebenen Hilfsmitteln angefertigt habe.

Yi, Dong
Wilhelm-Jäger-Str. 10
64287, Darmstadt

31.10.2012

# Erklärung

Ich erkläre hiermit, noch keinen Promotionsversuch unternommen zu haben.